STRUCTURAL STUDIES OF THE Ro RIBONUCLEOPROTEIN AND THE METALLOREGULATOR CsoR

A Dissertation

by

ARATI RAMESH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2007

Major Subject: Biochemistry

STRUCTURAL STUDIES OF THE Ro RIBONUCLEOPROTEIN AND THE METALLOREGULATOR CsoR

A Dissertation

by

ARATI RAMESH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James C. Sacchettini
Committee Members,	David P. Giedroc
	Patricia J. LiWang
	James C. Hu
Head of Department,	Gregory D. Reinhart

August 2007

Major Subject: Biochemistry

ABSTRACT

Structural Studies of the Ro Ribonucleoprotein and the Metalloregulator CsoR.
(August 2007)
Arati Ramesh, B.S., St. Ann's College, Hyderabad;
M.S., University of Pune
Chair of Advisory committee: Dr. James C. Sacchettini

Ro ribonucleoproteins are antigenic protein-RNA particles that are the major targets of the immune reaction in autoimmune disorders like systemic lupus erythematosus. The Ro protein has been implicated in cellular RNA quality control, due to its preference for binding misfolded non-coding RNAs such as pre5S ribosomal RNAs and U2 small-nuclear RNAs besides binding cytoplasmic RNAs called Y RNAs. Although well characterized in eukaryotes, an understanding of Ro in prokaryotes is lacking. To gain structural insight into Ro-RNA interactions we have determined a high resolution crystal structure of Rsr, a Ro ortholog from the bacterium *D. radiodurans.* The structure of Rsr reveals two domains- a flexible, RNA binding HEAT repeat domain and a cation binding vonWillebrand factor A domain. Structural differences between Rsr and *Xenopus laevis* Ro at the misfolded non-coding RNA binding site suggest a possible conformational switch in Ro that might enable RNA binding. Structural and biochemical characterization reveals that Ro binds cytoplasmic small RNAs called Y RNAs with low nanomolar affinity, to form ~700kDa multimers. Formation of these multimers suggests one possible mode by which Ro RNAs may be targeted towards downstream processing events.

Metal responsive transcriptional regulators sense specific metals in the cells and regulate the expression of specific operons involved in export, import or sequestration of the metal. CsoR is a copper(I) specific transcriptional regulator of the *cso* operon which

consists of a putative copper export pump, CtpV. In copper limiting conditions, CsoR binds the operator/promoter region of the *cso* operon. In increased concentrations of copper (I), CsoR binds copper (I) with high affinity and is released from the operator/promoter site, causing derepression of the *cso* operon. To gain structural insight into CsoR function, we have solved the crystal structure of copper(I) bound CsoR. The structure reveals a homodimer with a subunit bridging copper site. The trigonal planar geometry and the presence of cysteine and histidine ligands at the metal site are favorable for copper(I) binding. The structure reveals a novel DNA binding fold in CsoR, making it the founding member of a new structural class of metalloregulators.

For
amma, appa and *Preeti.*

ACKNOWLEDGEMENTS

First of all I would like to express gratitude to my advisor Dr. Jim Sacchettini for his constant guidance and encouragement throughout these years. His unique style of crystallography, mixing an old school approach with modern tools and programs has been an inspiration. I consider it my privilege to have learnt crystallography from him.

I thank all my committee members – Dr. Patti Liwang, Dr. Jim Hu, Dr. Andreas Baumler and Dr. David Giedroc for their helpful suggestions and scientific advice.

Many thanks to my friends Peter Cornish, Mario Pennella and Suzanne Stammler. Without their help, the RNA and fluorescence experiments could not have been performed as easily. Thanks to members of the Sacchettini group especially Dong Yang and Sudharsan Sridharan, who have patiently trained me in many of the crystallographic packages.

Finally, most of all I would like to thank my closest friend Christos Savva and my husband Sunil not only for scientific discussions and help with my experiments, but for their constant support and encouragement.

Thank you.

NOMENCLATURE

ATP	Adenosine 5'-triphosphate
APS	Advanced Photon Source
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dNTP	Deoxyribonucleotide triphosphate
dsDNA	double stranded Deoxyribo nucleic acid
dsRNA	double stranded ribonucleic acid
EM	Electron microscopy
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	isopropyl β-D- thiogalactopyranoside
MAD	Multiwavelength anomalous dispersion
β-ME	β-mercaptoethanol
MIR	Multiple isomorphous replacement
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
Se-Met	selenium-methionine
ssDNA	single stranded deoxyribonucleic acid
ssRNA	single stranded ribonucleic acid

Tris	2-amino-2-hydroxymethyl-1,3-propanediol
X	Any amino acid

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I
INTRODUCTION: STRUCTURAL BIOLOGY OF NUCLEIC ACID BINDING PROTEINS

1.1 INTRODUCTION TO STRUCTURAL BIOLOGY

Proteins, nucleic acids, lipids and other macromolecules form the basic building blocks of all cellular life forms. Structural biology aims to understand the architecture of these building blocks and the spatial arrangement of atoms in the macromolecules encountered in living organisms. The underlying molecular mechanisms of cellular processes cannot be fully understood without a thorough knowledge of the molecular architecture of these biological macromolecules.

The roots of macromolecular structural studies can be traced as far back as the middle of the 20th century. Most of the early work focused on X-ray structure determination of enzymes (1) and proteins that bind relatively simpler ligands (2). Pioneering work using X-ray crystallography in the 1950s, helped us understand the structure of the DNA double helix (3). Structural studies since then, along with biochemical and genetic studies have contributed greatly to our understanding of enzyme function, metabolic pathways and many other biological processes. Application of structure based studies has given rise to the relatively new field of rational drug design where macromolecular structures are used as a starting point for the design of inhibitors. This has been particularly helpful in understanding disease causing organisms and developing new drugs against them.

With advancement in technology and the emergence of newer techniques like nuclear magnetic resonance (NMR), electron microscopy (EM) and fluorescence, our understanding of macromolecular structure has multiplied exponentially. There are currently 39,815 structures available in the protein data bank (PDB, www.rcsb.org). Structural studies today address complex problems that range from understanding the

---

This dissertation follows the style of *Journal of Biological Chemistry*.

architecture and mechanisms of membrane bound proteins to large macromolecular complexes that could involve many different proteins.

Protein-nucleic acid interactions play critical roles in a variety of cellular processes. Fundamental processes like transcription, translation and regulation of gene expression all involve interactions between proteins and nucleic acids (4), (5). The understanding of these processes has been tremendously aided by the numerous structural biology approaches, which have provided visual, multidimensional snapshots into their functional mechanisms. Of the currently known structures, at least 1,751 are of protein nucleic-acid complexes (www.rcsb.org). From these studies, some common themes have emerged, that highlight the architectural features that proteins possess in order to interact favorably with nucleic-acids. Some of these common themes and structural motifs are discussed here.

1.2 STRUCTURAL STUDIES ON DNA-PROTEIN COMPLEXES

Double stranded DNA has a relatively uniform structure. The sugar-phosphate backbone is highly negatively charged and the core consists of base-pairs that stack against each other (6). At the major and minor grooves of the DNA, these stacked bases are exposed. The sequence of the bases along the DNA dictates the chemistry of the DNA surface. When transcription factors need to bind a specific sequence of DNA, it is this chemical surface that they need to recognize (7). Interactions between protein and DNA could be in the form of hydrogen bonds, van der Waal's interactions, salt-bridge interactions and stacking interactions (7). Hydrogen bonds could involve both side-chains and backbone of the protein and the bases as well as the backbone of the DNA (8). Multiple hydrogen bond interactions as seen in the Arginine-Guanine pair (Nδ and the Nε of the guanidinium group hydrogen bonding with guanine in a bidentate manner) and the Asparagine-Adenine or Glutamine-Adenine pair contribute to the specificity of the interaction (8). van der Waal's interactions give rise to steric constraints which further add to the specificity of interaction. Salt-bridges formed between the phosphate

backbone of the DNA and basic residues from the protein have also been shown to stabilize protein-DNA interaction (8). Stacking of residues like tryptophan and phenylalanine between base pairs has been shown in some cases (8), as discussed below.

Proteins that regulate transcription must possess the ability to respond to the effector molecule and in turn modulate their interactions with a specific DNA sequence. Structural studies in the past few decades have increased our understanding of at least some mechanisms governing such interactions. Structural studies have shown that transcriptional regulators could interact with DNA in a variety of ways, using either α-helices, β-sheets or loops (7). Commonly found DNA binding motifs are discussed here.

The *helix-turn-helix (HTH)* motif consists of two short helices at a ~120º angle, separated typically by a tight four-residue turn (9), (10). The second helix, called the recognition helix inserts into the DNA major groove to contact the base pairs as well as the backbone phosphates. High resolution structures of the bacteriophage lambda repressor/operator complex (PDB ID 1LMB, **Figure 1.1a**) show the interaction between the HTH motif and DNA (9). In the lambda repressor, in addition to the HTH motif, an N-terminal arm makes contacts with the DNA backbone. The angle at which the recognition helix interacts with the DNA differs between different protein-DNA complexes (11), with the maximum interaction between protein and DNA bases occurring when the axis of the inserted helix is parallel to the phosphate backbone (11). This type of interaction can be seen in the three-helix family of DNA and RNA binding proteins like the Mat α homeodomain proteins in yeast (12), **Figure 1.1b**. The DNA bound to such domains typically does not undergo distortion (7).

The *winged helix* motif is a variation of the helix-turn-helix motif and usually has an H_1-B_1-H_2-H_3-W_1-B_2-B_3-W_2 topology (13), **Figure 1.1c**. Helices H_2 and H_3 resemble an HTH motif with the H_3 being the recognition helix that interacts with DNA. Two wing-like loops (W_1 and W_2) are present on either side of the recognition helix, of which at least one (W_1) has been shown to make additional contacts with the DNA (13). A crystal structure of the mouse GA binding protein in complex with the GGA motif (PDB ID 1AWC) shows the details of interactions between a winged helix domain and the major groove of DNA (**Figure 1.1c**). The winged helix domain has been found in many families of metal responsive transcriptional regulators, for example the copper sensing repressor CueR from *E. coli* and the iron sensing IdeR protein from *M. tuberculosis* (14), (15).

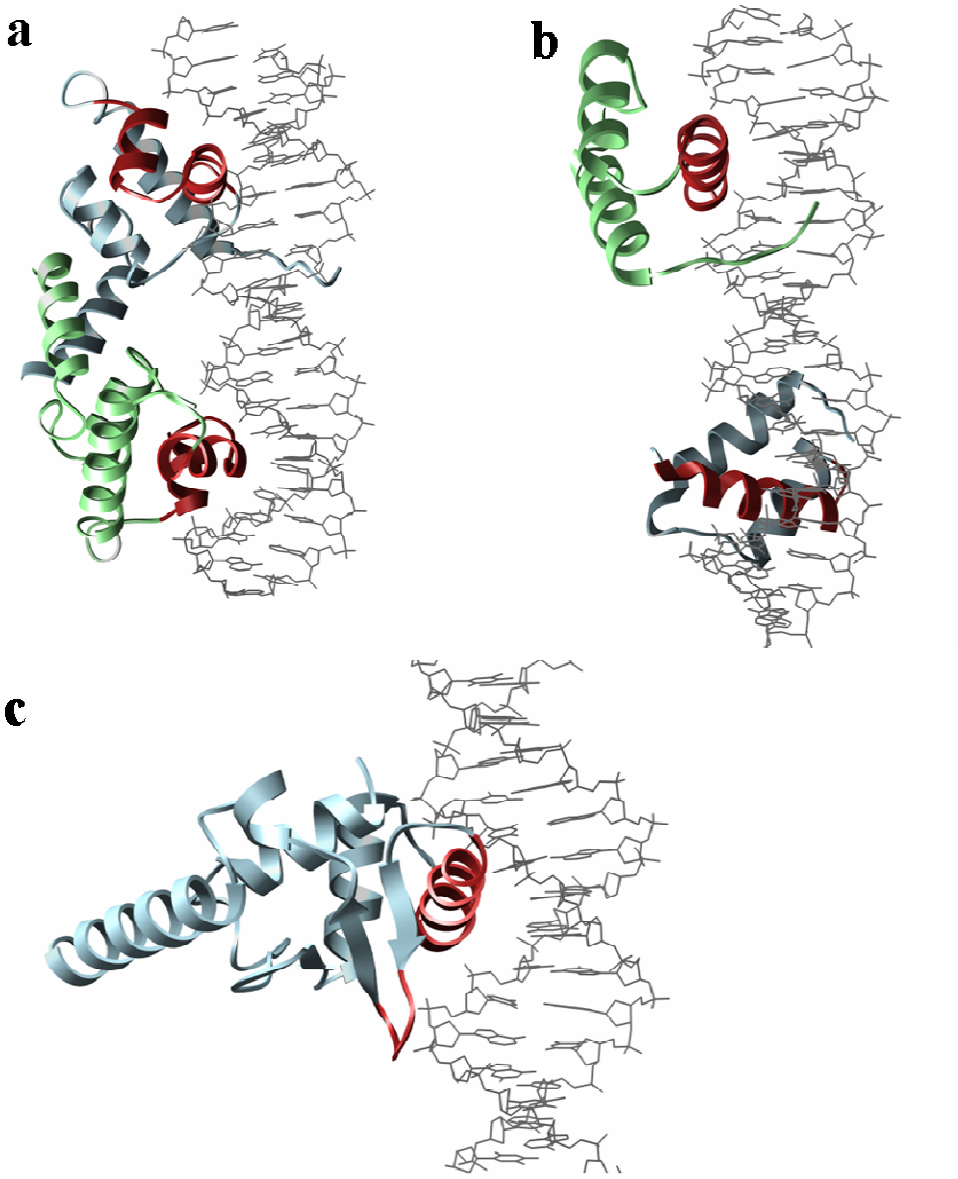


Figure 1.1 Helix-turn-helix motifs to bind DNA

a. Crystal structure of lambda repressor /operator complex (1LMB) with a helix-turn-helix motif (red). The recognition helix interacts with the major groove of the DNA (grey). Monomers are shown in blue and green. b. Crystal structure of a Mat-α2 homeodomain/operator complex (1APL) shows a three helical bundle with the recognition helix (red) inserted into the major groove of the DNA (grey). Monomers are shown in blue and green. c. Crystal structure of a domain of the transcriptional regulator GABP in complex with DNA (1AWC) shows a winged helix domain with the recognition helix and wing (red) interacting with the major groove of the DNA (grey).

Other helical motifs that interact with the major groove of the DNA are the *basic region leucine zipper (bzip)* motif and the *helix-loop-helix (HLH)* motif (16), (17). These motifs are found in dimeric eukaryotic proteins (7). The bzip motif typically consists of long, ~60 residue α-helices that form coiled coils (**Figure 1.2a**). A structure of the human proto-oncogene product C-Jun in complex with the cyclic AMP recognizing element (CRE) DNA shows interactions between a bzip motif and the major groove of DNA (PDB ID 1JNM, **Figure 1.2a**). The C-terminal dimer interface is stabilized by hydrophobic interactions between leucine residues. The N-terminal portion of the α-helix is inserted into the major groove of the DNA (7). The HLH family of proteins also dimerizes, but each monomer in the symmetric homodimer is composed of two α-helices joined by a loop so that the dimer interface consists of a four-helix bundle (17). The structure of the mouse transcription factor Max, in complex with its cognate DNA (PDB ID 1AN2, **Figure 1.2b**) shows that the HLH motif interacts with DNA in the same way as the bzip motif.

Proteins like the *Lac and purine repressors* from *E. coli* interact with the minor groove of the DNA (18) (19), although base stacks are not optimally exposed in the minor groove (6). To accommodate the protein side chains in the minor groove, the DNA undergoes sufficient distortion. For example, the Purine repressor PurR binds DNA, causing a 45º kink, unwinding and base unstacking in the DNA (PDB ID 1PNR, **Figure 1.3**). The protein forms a dimer with helices from each monomer coming together to interact with the minor groove. Leucine residues from the protein are intercalated into the base-stack to facilitate the kink in the DNA (19). Each monomer also possesses a separate helix-turn-helix motif forming a 'head' which interacts with the neighboring major groove (19).

Helical DNA binding folds in proteins could also be stabilized by metal coordination in close proximity to the DNA binding site (20), (21). The *zinc finger motif* proteins contain multiple modules of an ~ 30 residue DNA binding domain (**Figure 1.4**) (20). A crystal structure of the mouse Zif268 transcription factor in complex with DNA

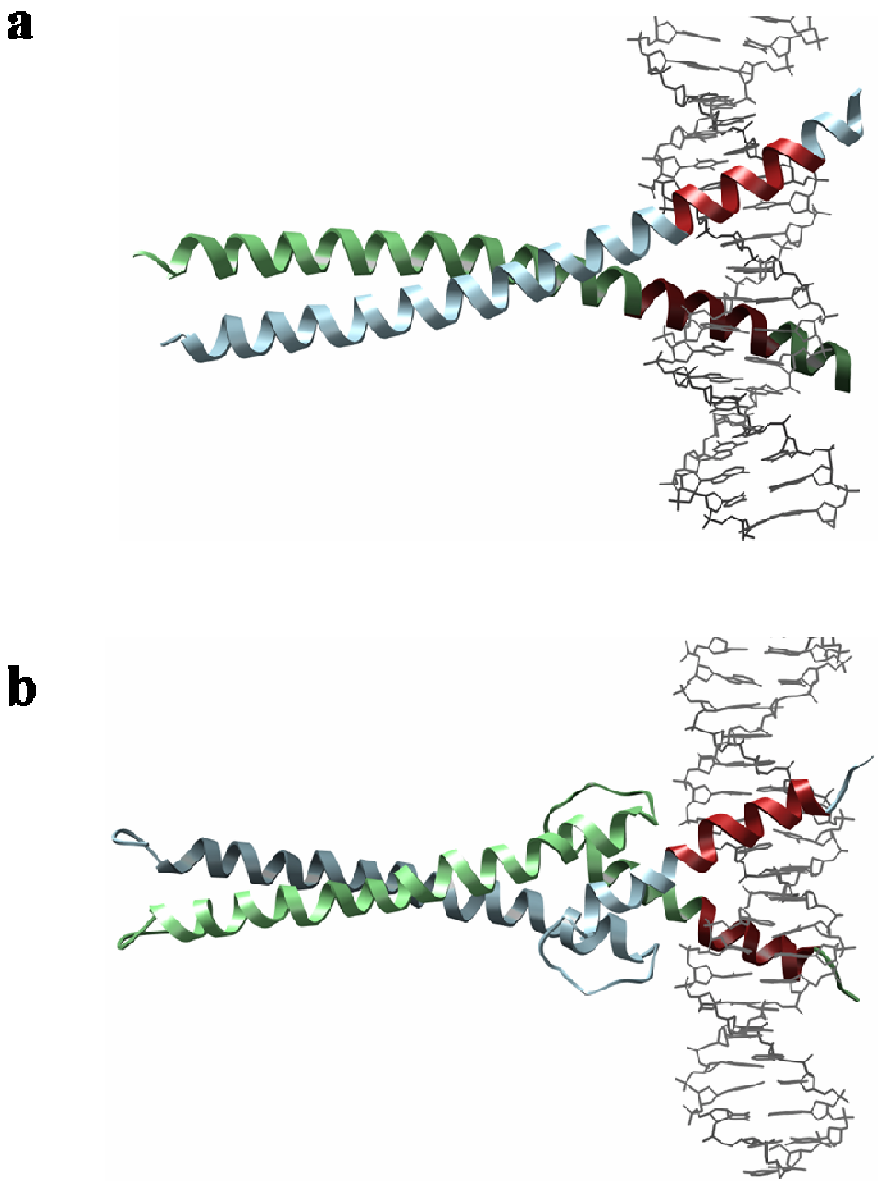


Figure 1.2 α-helical domains to bind DNA

a. Crystal structure of the Jun/Cre complex (1JNM) shows a basic region leucine zipper (bzip) motif. Coiled coils formed by two monomers (blue and green) are stabilized by a leucine zipper core. The C-terminal portion of the helices (red) interacts with the major groove of the DNA (grey). b. Crystal structure of the transcription factor Max in complex with DNA (1AN2) shows a dimer (blue and green) with each monomer consisting of a helix-loop-helix motif. The dimerization interface resembles the leucine zipper and the ends of the helices (red) insert into the major groove of the DNA (grey).

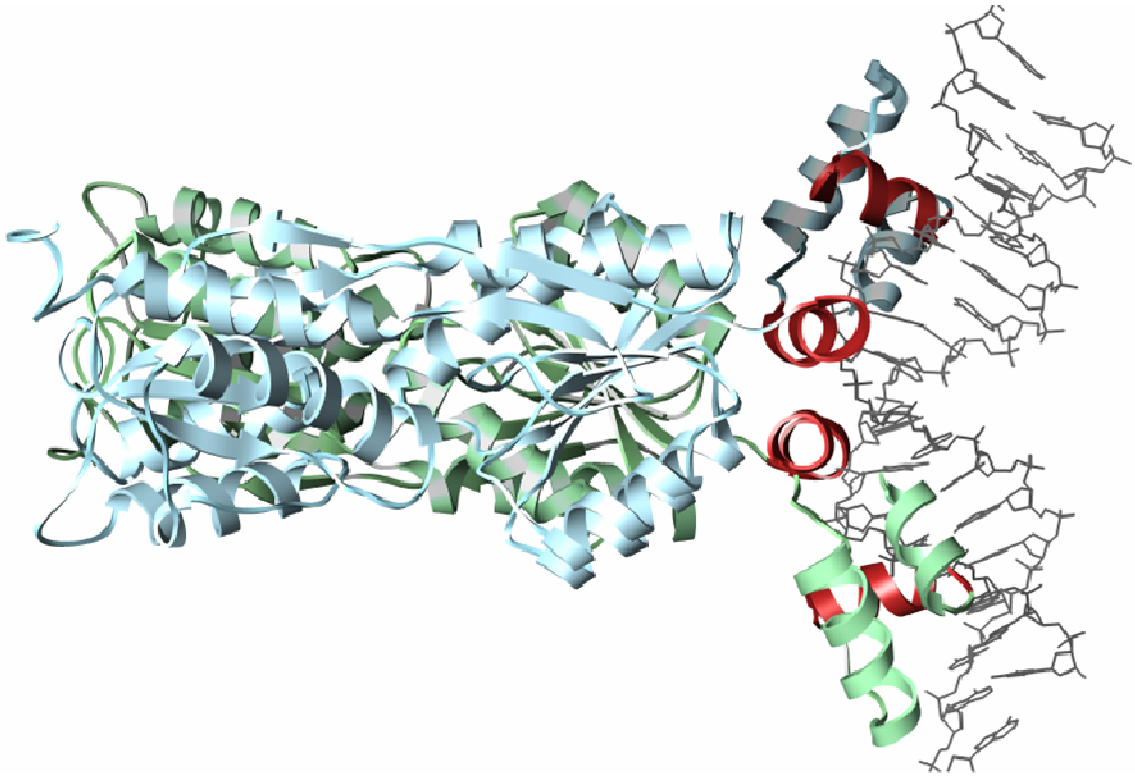


Figure 1.3 α-helical domains to bind minor groove of DNA

Crystal structure of the purine repressor PurR bound to DNA (1PNR) shows a dimer (blue and green) with recognition helices (red) from each monomer coming together to bind the minor groove of the DNA (grey). To accommodate the helices at the minor groove, the DNA undergoes a huge distortion with a kink and unwinding. Helices from each monomer (red) separately contact the major grooves too.

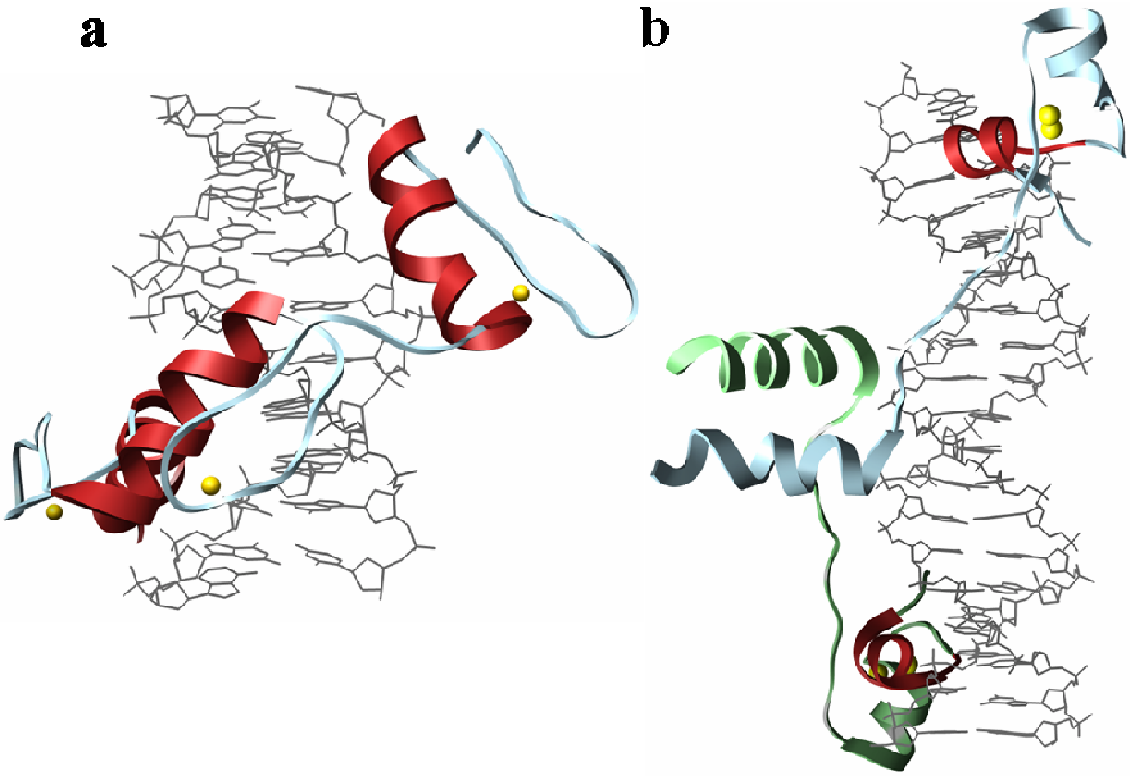


Figure 1.4 Zinc coordinating motifs to bind DNA

a. Crystal structure of Zif268/DNA complex (1AAY) shows multiple modules of zinc finger motifs. The recognition helix (red) from each module interacts with subsequent major grooves of the DNA (grey). The zinc finger itself is stabilized by a zinc ion (yellow). b. Crystal structure of the yeast transcriptional factor Gal4 in complex with DNA (1D66) shows a homodimer (blue and green). The dimerization domain interacts with the minor groove of the DNA (grey) and recognition helices (red) from each monomer interact with the different major grooves.

While the majority of DNA-binding motifs are α-helical, DNA recognition also takes place through β-sheets. Proteins with the *ribbon-helix-helix* motif form dimers with a β-strand from each monomer coming together to form a two stranded antiparallel β-sheet (**Figure 1.5**) (22), (23). The crystal structure of the omega repressor from *S. pyrogenes* (PDB ID 2CAX, **Figure 1.5**) shows that a β-sheet inserts into the major groove of the DNA such that the plane of the sheet is parallel to the DNA backbone (22). The side chains normal to the plane of the β-sheet interact with the base-pairs (22). The Arc repressor of bacteriophage P22 forms a dimer of dimers with each protomer contributing a β-strand that interacts with the DNA major groove (24). Similarly, a nickel responsive transcriptional regulator NikR, from *E. coli* recognizes DNA through a ribbon-helix-helix motif (23). Recognition of DNA through a β-sheet also occurs at the minor groove as seen for TATA binding proteins (TBP) (25). The yeast TBP inserts a ten stranded β-sheet into the minor groove of the TATA box DNA, causing huge distortion of the DNA, along with unwinding and bending (PDB ID 1YTB, Figure 1.5). The bending and distortion of the DNA is facilitated by phenylalanine side chains that are intercalated into the DNA base-pair stacks (25).

One family of DNA binding proteins recognizes DNA through loop regions (26). This family of p53-like transcription factors consists of a β-sheet immunoglobulin-like domain. The crystal structure of human NFκB p53 homodimer in complex with DNA

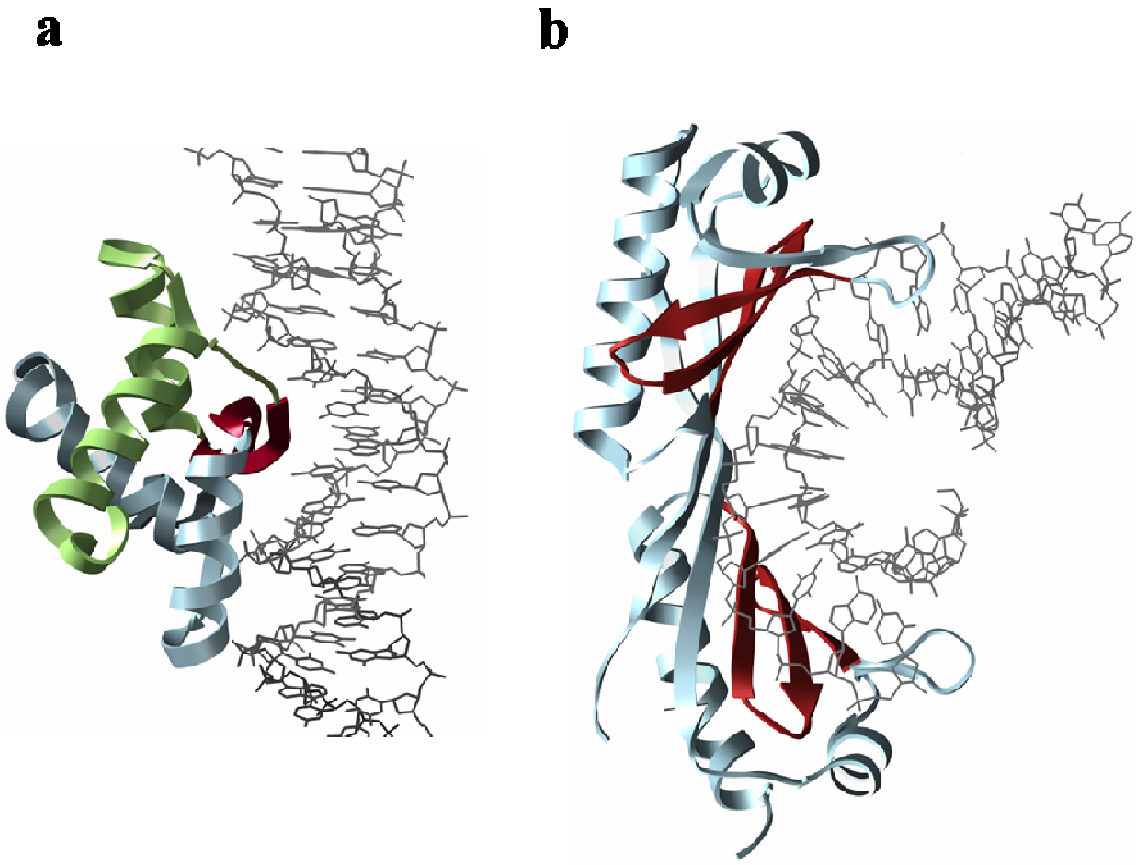


Figure 1.5 β-sheet motifs to bind DNA

a. Crystal structure of the omega repressor/DNA complex (2CAX) shows 2 β-strands (red) from each monomer (blue and green) coming together to form a β-sheet. The β-sheet interacts with the major groove of the DNA (grey). b. Crystal structure of the yeast TATA binding protein in complex with the TATA box (1YTB) shows parts of a 10 stranded β-sheet (red) interacting with the minor groove of the DNA (grey). The DNA undergoes a kink with unwinding and base unstacking.

shows interactions via side-chains at the dimerization interface of the protein (PDB ID 1A3Q, **Figure 1.6**). The α-helical region is rotated away from the dimerization interface, leaving loop regions to interact with the major groove of the DNA (26).

Prokaryotic transcriptional regulators belong to the helix-turn-helix, winged helix and β-ribbon families, barring few exceptions. They bind to cognate DNA sites that are almost invariably palindromic or pseudopalindromic (7).

1.3 STRUCTURAL STUDIES OF PROTEIN-RNA COMPLEXES

RNAs can adopt a variety of structures in single-stranded as well as double stranded conformations. Secondary structural elements like hairpins, loops, single or more nucleotide bulges are commonly found in RNAs (27), (28). Double stranded RNA helices are found in the A form in which the centre of the base pair is displaced by ~4Å as compared to the classical B-form helix (29). As a consequence, the minor groove is wide (~11Å) and shallow (~3Å) in the A form and the major groove very narrow (~3Å) and deep (~13Å), in contrast to the wide and less deep major groove in B-form helices (29). This classical picture has been confirmed in investigations of RNA double helical tracts by crystallography (30) and NMR (31). While proteins interacting with DNA often do so by forming specific hydrogen bonds between the bases and the side chains of an alpha-helix inserted into the accessible major groove of the B-form DNA, the much narrower and deeper groove of the A-form RNA helix is less accessible to the α-helices of RNA binding proteins. Often, unpaired nucleotides stacked inside an RNA helix or bulged outside the helix represent inviting targets for site-specific recognition of RNA by proteins (32).

RNAs play a variety of roles in both eukaryotic and prokaryotic cells (33), (34), (35). Besides the messenger RNAs (mRNA) that are involved in message transfer, a host of non-coding RNAs like ribosomal RNAs (rRNA), transfer RNAs (tRNA), small nuclear RNAs (snRNA), small nucleolar RNAs (snoRNA), micro RNAs (miRNA), anti-sense RNAs etc. have been discovered (33). The various roles played by these RNAs

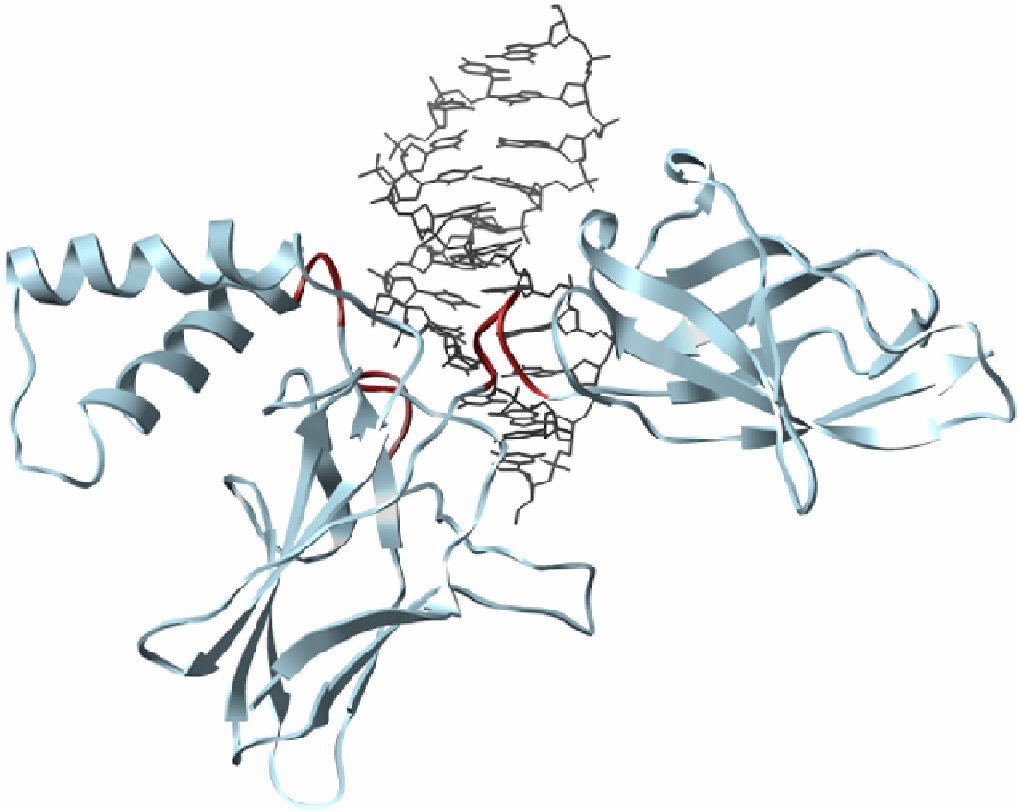


Figure 1.6 Loop regions to bind DNA

Crystal structure of human NF-kB bound to DNA (1A3Q) shows major interactions between loop regions (red) and the major groove of the DNA (grey).

range from being structural molecules to splicing and gene regulation. As expected, proteins that interact with RNAs are involved in a huge variety of cellular processes. As a result, proteins may need to distinguish between RNA and DNA, single stranded and double stranded RNA and even different structural elements like hairpin loops, bulges and pseudoknots.

The interaction of RNA binding proteins with RNA has been shown to involve a wide variety of very elegant structural features in both the protein as well as the RNA and these features often vary from subtle to the obvious. High resolution structure determination by X ray Crystallography and Nuclear Magnetic Resonance (NMR), two of the most comprehensive techniques used to study the structure of RNAs and RNA binding proteins, has made it possible to understand structural phenomena in a very elaborate and accurate manner. Several different structural motifs have been identified in RNA binding proteins (32).

The most commonly found and most well studied RNA-binding motif is the βαββαβ - RNP consensus sequence (RNP-CS) RNA-binding domain (RBD) (32) (**Figure 1.7**). The motif comprises of 90-100 amino acids and the consensus sequence (also called RNA recognition motif or RRM) is evolutionarily conserved. Proteins with this domain localize to the nucleus, nucleolus, cytoplasm or cytoplasmic organelles, interact with a variety of RNAs like pre-mRNA, pre-rRNA, snRNA and mRNA and comprise the largest family of RNA-binding proteins (36). The RNP-CS can be identified in the amino acid sequence of a protein by the presence of the highly conserved RNP1 octamer (K/R G F/Y G/A F V X F/Y) and RNP2 hexamer (L/I F/Y V/I G/K N/G L) sequences which have an array of aromatic residues and a number of hydrophobic residues (36). The two RNP sequences are separated by ~ 30 residues. RBDs are often present as multiple modules in proteins, possibly to increase the specificity of interaction by creating multiple interacting surfaces (36). Modules of RBDs are typically connected by a 10 to 15 amino-acid linker region. This linker region itself has been shown to interact with RNA, often being unstructured in the apo form of the protein and becoming ordered in the RNA bound form (32), (37). Primarily,

interactions between RNA and the RRM proteins occur via the insertion of multiple loops into the minor groove of RNA (**Figure 1.7**). Additional interactions come from the side-chains on the face of the β-sheet with the RNA being placed along the face of the β-sheet. The number of β-strands that make up the β-sheet varies. In the human U1 snRNP A protein, an antiparallel four-stranded β sheet packs against two perpendicularly oriented α helices (38). The RNP 1 in the β3 segment and the RNP 2 in the β1 segment are juxtaposed to create an interface packed with aromatic residues. Positively charged residues are located on the loops at one end of the β-sheet and mutations in this region have a significant effect on RNA binding. As seen in most other RRM proteins, the linker region between the two RNPs is highly disordered in the crystal structure of the apo-protein. The structure of this domain in complex with a 21 nucleotide synthetic RNA representing the U1 snRNA hairpin (39) shows that the most significant differences between the structures in the apo and complex state, are in the loops connecting β2-β3 and helix2-β4 (39). The linker region gets ordered when bound to RNA. The 10 nucleotide loop of the RNA fits into the groove between the β2-β3 loops and interacts with the RNP sequences as a single stranded RNA (39). The RNA binds the β sheet in a large area which is more exposed and open like a platform rather than a buried niche. The arginine residues in the RNP motifs form hydrogen bonding interactions with the RNA base pairs. Various stacking interactions between nucleotide bases and aromatic residues from the protein additionally stabilize the complex. The residues that stack with the RNA bases typically are held in place by hydrogen bonds between their backbones. The sugar-phosphate backbone of the RNA interacts with the positively charged surface of the protein. Hydrogen bonds between backbones of residues position the aromatic residues to stack against pyrimidine rings. This possibly decreases the flexibility of the large rings and tightens the RNA protein interaction. This probably also confers specificity to the interaction since it involves side-chains and not backbone interactions. Despite the common structural fold of RBD proteins, they bind a host of different RNAs. Information about specificity seems to lie in the residues just before the RNP2 and just after the RNP1, in the loop regions extending beyond the

aromatic residues which make contacts with the side-chains.

Another commonly found RNA binding motif is the αββαβ- K homology (KH) domain (32). This domain is characterized by a conserved octapeptide sequence- Ile-Gly-X_2-Gly-X_2-Ile. Many repeats of this sequence are often present in the protein. Although initially identified in hnRNP K, the domain was found to be prevalent in numerous other RNA associated proteins like Mer 1p, the ribosomal S3 proteins and the yeast alternative splicing factor (40), (32). The KH domain consists of a three stranded β-sheet exposed on one surface and two α-helices on the opposite side forming a helix turn helix motif (41). The KH domain assumes the same overall topology as the RBD with minor differences. However, the mode of interaction with RNAs is very different in these two domains. RNAs interact with the KH domain not along the face of the β-sheet as seen in RBDs but on the helical side of the domain (32), (**Figure 1.7**). The short loop between the two helices has been shown to make critical contacts with the RNA and often goes from an unstructured to a more ordered form upon RNA binding. A structure of KH domain from hnRNP K in complex with ssDNA (42) shows that besides RNAs, this domain could also interact with DNA via the helix turn helix motif. This interaction involves both the backbone and the bases of the DNA. Base contacts are made both by stacking against aromatic residues as well as by hydrogen bonding. Although there seems to be a common theme in the way the RBD and KH domains interact with RNAs, individual proteins of conserved families have been seen to develop special interactions with their substrates (37). This may provide another determinant of specificity.

Yet another commonly found RNA binding domain is the all helical Pumilio Homology Domain (PUM HD) or Puf Domain. The Puf domain is named after the members of the Puf family of proteins which regulate expression of mRNA expression by binding the 3′ untranslated regions (UTRs) of their mRNA targets (43). This RNA binding domain consists of eight to nine α-helical repeats. The substrate RNAs typically contain two conserved boxes Box A (GU<u>UGU</u>) and Box B (AU<u>UGU</u>A). The UGU sequence has been shown to be essential for interactions with the protein. The structure

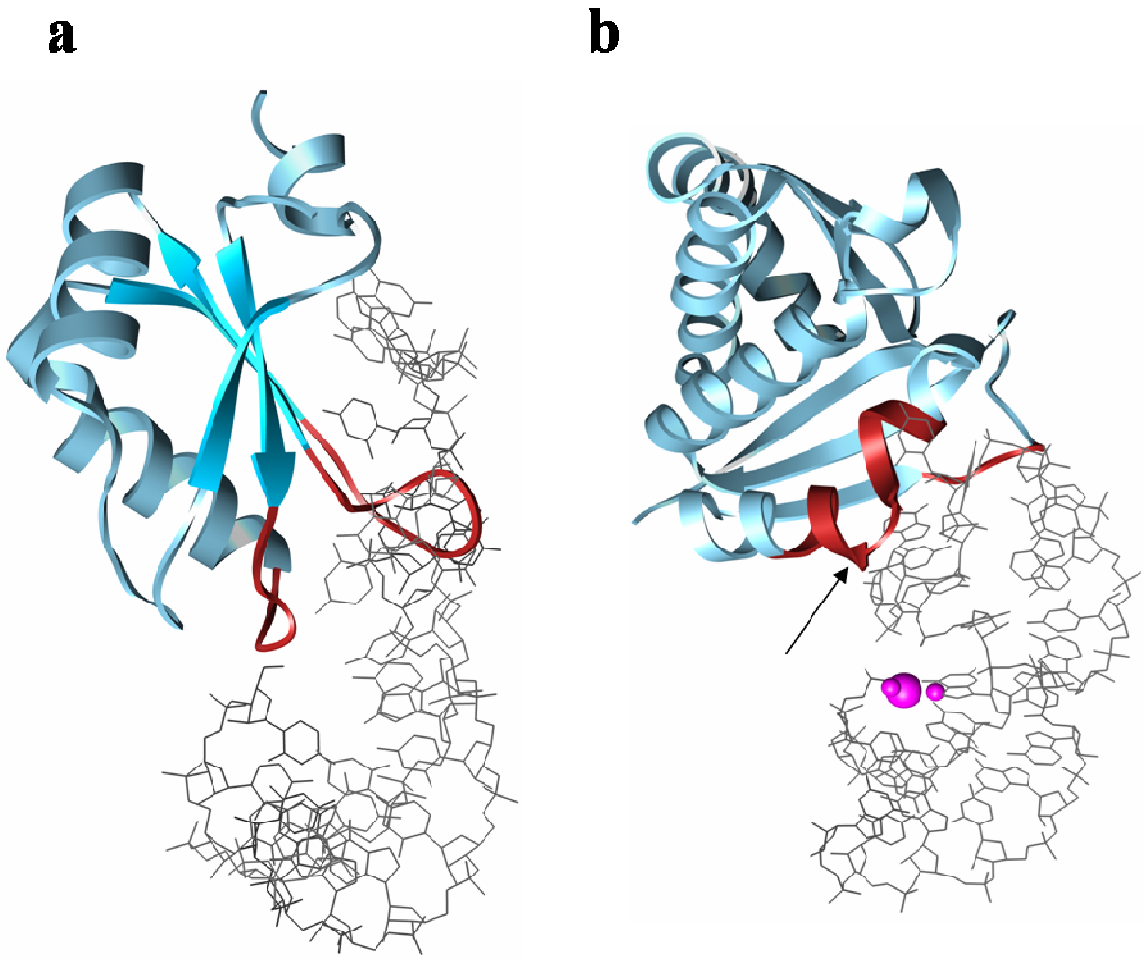


Figure 1.7 The RNA recognition motif (RRM) and the K-homology domain

a. Crystal structure of spliceosomal U2A-U2B in complex with U2 snRNA (1A9N) is shown. A 4-stranded antiparallel β-sheet forms part of the RNA (shown in grey) binding surface. Loops (red) between β-strands and a β-strand and α-helix insert into the minor groove of the RNA. b. Crystal structure of the KH domain of neuronal splicing factor Nova1 in complex with a 25nt. RNA hairpin (2ANN) is shown. The RNA (grey) interacts with the helical side (red) of the KH domain. The loop (black arrow) connecting the two helices becomes ordered upon binding RNA.

of HsPUM HD bound to a 10mer NRE RNA fragment (43) shows that the interactions between the PUM HD and the RNA take place along an extended, concave surface (**Figure 1.8**). The bases of the RNA face the protein interface with the RNA backbone facing away from the protein. Bases from the RNA are involved in stacking interactions with residues from the α-helical repeats of the protein. Hydrogen bonding interactions are present between the bases and residues at the concave interface, in addition to multiple van der Waal's contacts that further stabilize the interaction. Superposition of the backbone atoms of the protein in the apo and RNA bound form do not suggest any significant conformational changes in the protein upon binding RNA. The structural studies reveal the possible determinant of specificity for this domain, since majority of contacts are made with the bases of the RNA and not with the sugar-phosphate backbone. In some respects, this domain resembles the TROVE domain found in the Telomerase, Ro and Vault RNPs (44). The TROVE domain itself is a ~ 300 amino acids long domain with multiple short helices connected by short loops. Some TROVE domain proteins also have the RNP consensus sequences characteristic of the RRM, though the role of the RNPs in these domains is not clear (44).

Double stranded RNA-binding proteins (DRBPs) form a class of proteins with specificity towards double-stranded RNAs (45). These proteins contain an αββββα type double-stranded RNA binding domain (DRBD) characterized by the presence of an evolutionarily conserved motif (46). Proteins of this family may contain up to five DRBDs, each being ~ 65 to 68 amino-acids long (45). The crystal structure of the dsRBD from *Xenopus laevis* dsRNA binding protein A in complex with dsRNA (PDB ID 1DI2) shows that three β-strands form an antiparallel β-sheet against which the two α-helices are packed (47), (**Figure 1.9**). Majority of the contacts with RNA involve 2'OH and phosphate groups instead of the sequence specific interactions with the bases (47). A structure of the Drosophilla Stauffen DRBD in complex with an RNA loop (48), (PDB ID 1EKZ) shows how proteins belonging to this family could not only interact with dsRNA but also with ssRNA with elaborate secondary structure elements, but not with ssDNA, dsDNA and ssRNA.

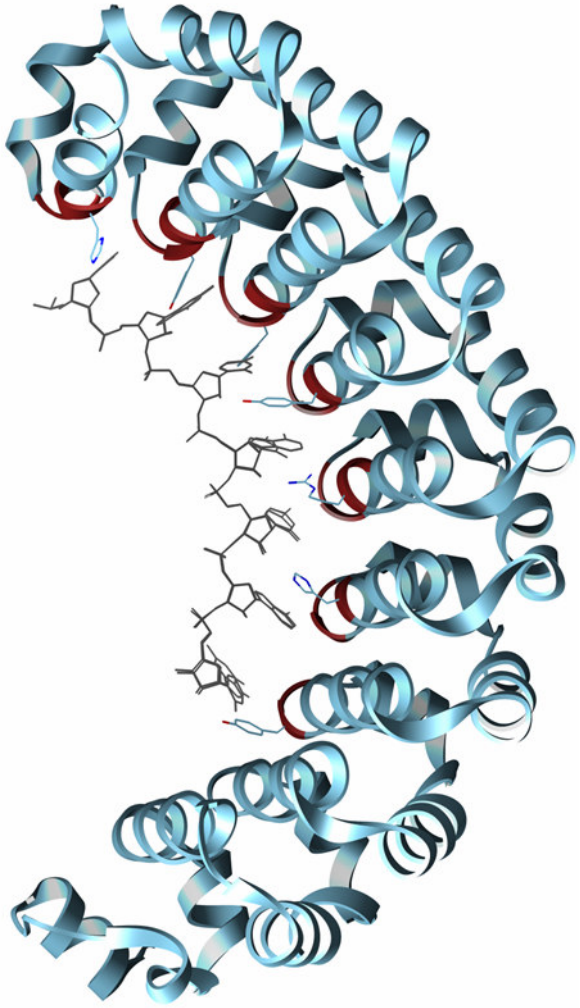


Figure 1.8 The RNA binding Pumilio homology domain

Crystal structure of the human pumilio homology domain in complex with the NRE1-19 RNA (1M8W) is shown. The extensive RNA binding interface (red) is concave and lined with aromatic residues and charged residues. The aromatic residues are involved in stacking against RNA bases (grey). Electrostatic and van der Waal's interactions further stabilize the complex.

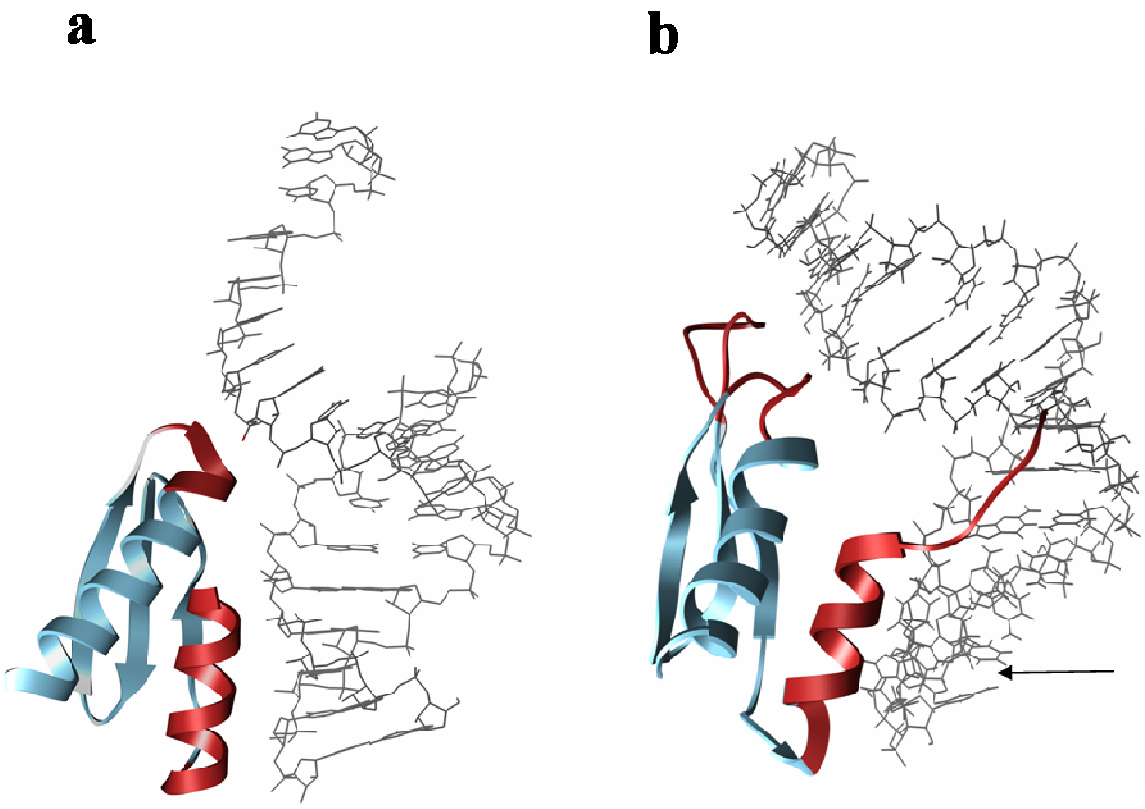


Figure 1.9 The double stranded RNA binding domain (dsRBD)

a. Crystal structure of dsRNA binding protein A from *X. laevis,* in complex with dsRNA (1DI2) is shown. The protein forms antiparallel β-sheet with the α-helices interacting with primarily the backbone of the RNA (grey). b. NMR structure of the third dsRBD from drosophila Staufen in complex with an RNA hairpin (1EKZ) is shown. The dsRBD interacts with the minor groove and backbone of dsRNA (grey) via conserved residues in two loops (red). An α-helical region (red) additionally interacts with the single stranded hairpin region of the RNA (black arrow).

1.4 OVERVIEW OF DISSERTATION CHAPTERS

With Chapter I as an introduction to the general themes in protein nucleic-acid interactions, in the following chapters the structural studies performed on two nucleic acid binding proteins are described. To obtain high resolution structural information, X-ray crystallography has been used. In chapter II the study of the RNA-binding Ro protein from *Deinococcus radiodurans* has been described. The Ro protein is involved in binding misfolded RNAs as well as a class of non-coding small cytoplasmic RNAs called the Y RNAs. There is growing evidence that Ro may play a role in the RNA quality control in cells. These structural studies elucidate the conformational changes that might take place in the Ro protein to facilitate RNA binding. This study also provides clues to the function of Ro in prokaryotes.

In chapter III, the study of a metal responsive transcriptional regulator, CsoR from *Mycobacterium tuberculosis* is described. CsoR binds to a specific operator/promoter region of DNA in its apo form. CsoR responds to free Cu(I) ions in the cell and causes derepression of the copper sensitive (*cso)* operon. In this way, it regulates the expression of a putative Cu(I) sensing P-type ATPase and other proteins. These structural studies reveal the mode of Cu(I) binding by CsoR and suggest possible determinants of Cu(I) specificity in CsoR. Structure based comparisons suggest that CsoR uses a novel structural fold to interact with DNA.

In both Chapters II and III, fundamentals of protein crystallography have been discussed, using examples from our studies on Ro and CsoR.

CHAPTER II

STRUCTURAL STUDY OF THE Ro PROTEIN

2.1 INTRODUCTION

2.1.1 Introduction to RoRNPs

RoRNPs are a class of RNA protein complexes found in most vertebrate cells. The major protein component of the RoRNPs has been shown to be a 60kDa protein named Ro (49). The RNA components of the Ro RNPs are a class of small cytoplasmic non-coding RNAs called Y RNAs, which are transcribed by RNA polymerase III and range in size from 69 to 112 nucleotides (50). RoRNPs were discovered more than four decades ago. Research since then has been directed towards understanding these RNPs, due to an intense medical interest in these RNPs.

2.1.2 Disease pathology associated with RoRNPs

Ro RNPs were first identified in patients suffering from autoimmune disorders like Systemic Lupus Erythematosus (SLE) and Sjögrens syndrome (SS) (51), (52), (53). Systemic autoimmune disorders like SLE are caused when the immune system wrongly recognizes the healthy cells and tissue in the body as “foreign” and targets them by producing autoantibodies. As seen for most autoimmune disorders, the damage is not limited to any particular part of the body, affecting the heart, kidneys, skin, joints etc. (54). Nephritis, pleuritis, arthritis and unusual sensitivity to light are common symptoms of lupus like autoimmune disorders. The sera of patients suffering from such autoimmune disorders have a higher than normal titre of autoantibodies which recognize Ro RNPs (53). Anti-Ro antibodies are also associated with neo-natal Lupus Erythematosus (NLE) in which trans-placental movement of antibodies takes place from mothers suffering from lupus to their babies (55). NLE is often associated with congenital heart block and other cardiac, hepatic and dermatological disorders. Lerner et al. have shown that the protein component of the Ro RNPs is required for the antigenicity since the Anti-Ro antibodies do not recognize the RNA components alone

(52). The immune response is heterogeneous and in some cases, the antibodies recognize only the native protein and not the denatured polypeptides (56). Currently, no cure is known for SLE and SS and the only treatment available is symptomatic, using immunosuppressants and corticosteroids (57). A better understanding of the antigenic RoRNPs has resulted from studies in other organisms.

2.1.3 Role of RoRNPs in eukaryotes

In eukaryotes, Ro and Y RNAs have been identified in all vertebrates and the nematode *Caenorhabditis elegans* (58), (59), (60). The number of distinct Y RNAs bound by Ro varies between different species. Humans and *X. laevis*, for example contain four known Y RNAs, mouse cells contain only two and *C. elegans* contains only one Y RNA (61), (58). Y RNAs in humans have been implicated in DNA replication (62). It has been shown that a subset of Ro RNPs also contains other protein components for example the 50kDa La protein which binds to newly synthesized polymerase III transcripts like a number of nascent small RNAs including Y RNAs (63). Another protein seen to associate with Ro is the 52kDa Ro protein, though the association between the 60kDa and 52kDa Ro has not been very well established (64), (65). Ro as well as Y RNAs have been found in the cytoplasm as well as the nucleus (66).

In mice, the absence of Ro leads to the development of a lupus-like syndrome, characterized by autoantibodies against chromatin and ribosomes and glomerulonephritis and photosensitivity (67) leading to the proposal that there may be a normal function of Ro that is important for the prevention of autoimmune reactions. Also, mouse embryonic stem cells lacking Ro are highly sensitive to ultraviolet radiation (68). Both Ro and Y RNA accumulate in the cells after UV irradiation , showing a drastic change in the cellular distribution from cytoplasmic to nuclear (68). In mouse embryonic cells, Ro also binds misfolded spliceosomal U2 small nuclear RNAs (68). These misfolded U2 snRNAs are characterized by the formation of an abnormal helix not found in the normal U2 snRNAs. This abnormal helix has been shown to be required for Ro binding by the misfolded U2 snRNAs (68). It is not yet known whether these misfolded U2 snRNAs

assemble into functional snRNPs that retain the normal splicing activity or get degraded before being assembled.

Another well characterized ortholog of Ro is from *Xenopus laevis*. In oocytes of *X. laevis*, Ro has been identified in complex with mutant 5S rRNAs in addition to the four Y RNAs found in *Xenopus* somatic cells (69). These mutant 5S rRNAs contain additional nucleotides at the 3' end, consistent with the failure of RNA polymerase III to terminate at the first transcription termination signal, and also contain one or more point mutations compared to the major oocyte 5S rRNA sequence (69). It has been shown that the mutations in the sequence of these mutant pre 5S rRNAs results in a structural change in the RNA, allowing the formation of an alternate structure which differs from the normal pre 5S rRNAs (70), (71). This helical region along with the 3' single stranded extension has been shown to interact with Ro (72). 5S rRNAs are small non-coding RNAs that from part of the large subunit of the ribosome and are essential for translation (73). *X. laevis* has ~20,000 genes that code for 5S rRNA, presumably for the large numbers of ribosomes required during the early developmental stages. These misfolded pre5S rRNAs recognized by Ro are unfit to follow the pathway to maturation and eventually get degraded (69), (74).

Misfolded RNAs bound by Ro contain a single-stranded 3' extension which has been shown to form part of the Ro binding site (72). Binding of Ro with these misfolded rRNAs and U2 snRNAs has led to the proposal that Ro may function as part of a novel quality control or discard pathway for misfolded RNAs. Interestingly, mice have been shown to have a large number of U2 snRNA genes just as *X. laevis* has an exceptionally large number of 5S rRNA genes. Based on this observation it has been proposed that a sudden, high abundance of RNAs in the cell could be associated with a requirement of the cell to deal with RNA folding problems (68). Ro may play a role by binding the misfolded RNAs and helping them either fold or get degraded, through a yet unknown mechanism. In *in vitro* experiments it has been seen that Ro also binds the normal pre5S rRNAs and U2 snRNAs (72), though in physiological conditions these RNAs are likely to be bound by other proteins (75), (76), (77), (78). Although reasonably well

characterized in eukaryotes, little is known about the role of Ro in prokaryotic organisms.

2.1.4 Role of RoRNPs in prokaryotes

Following the explosion of genome sequencing projects, orthologs of Ro have been found in the genome sequences of many organisms like the algae *Chlamydomonas reinhardtii*, the cyanobacteria *Nostoc punctiforme* and *Synechococcus*, the eubacteria *Deinococcus radiodurans* and *Mycobacterium smegmatis*, the planctomycete *Pirellula* and the mycobacteriophage Bxz1 (79), (80), (81). Although present in only few prokaryotes, the Ro protein is highly conserved in all the species where it has been identified so far, with ~35% sequence identity between any two Ros. The genome sequences of *Mycobacterium smegmatis,* a non-virulent strain of mycobacteria and Bxz1*,* a mycobacteriophage specific for *M. smegmatis* also contain an ortholog of Ro. Whether Y RNAs or Y RNA-like RNAs exist in *M. smegmatis* and Bxz1 is still a question and one of the goals of this project is to try to answer this question. Of all the Ro proteins from lower organisms, the Ro ortholog from *D. radiodurans,* called Rsr (Ro-sixty related) is the best characterized.

D. radiodurans is a polyextremophillic eubacterium which can survive unusual amounts of heat, cold, dehydration and radiation (82). *D. radiodurans* has been shown to tolerate close to 5000 Gy of UV radiation which is ~85 times that tolerable by *E. coli* and 500 times that tolerable by humans. It has been shown that the Rsr protein contributes to the resistance of *D.radiodurans* to UV irradiation (83). *D. radiodurans* cells lacking Rsr are more sensitive to UV radiation than wild-type cells and this phenotype can be partly complemented by Rsr expressed under the control of a heterologous promoter. Rsr binds several small RNAs encoded upstream of *rsr,* which accumulate following UV irradiation (83). These RNAs, of which at least one is predicted to assume a secondary structure that resembles a Y RNA, can be immunoprecipitated along with Rsr using anti-Rsr antibodies (83). Studies in the last decade have shown that the amazing ability of *D. radiodurans* to tolerate such high

doses of radiation could be attributed to a number of different strategies that it uses to deal with UV damaged DNA (82). *D. radiodurans* has a high genome copy number providing additional copies of essential genes that might get damaged. A tight ringlike nucleoid prevents the diffusion of damaged DNA, probably increasing the accessibility of damaged DNA to repair enzymes (84). Increased concentration of Manganese ions (Mn^{2+}) post UV irradiation possibly allows scavenging of harmful reactive oxygen species that accumulate (85). Besides these passive strategies, there are a variety of DNA repair enzymes that are triggered post UV damage. These are shown to be both RecA dependent and RecA independent repair processes like non-homologous end joining (NHEJ) and single strand annealing (SSA) (86), (87), (88))

Based on the studies in *D. radiodurans* and mouse embryonic stem cells it has been proposed that Ro RNPs might contribute to recovery of cells following UV irradiation by acting as a sink for damaged RNAs (68). The relation between the role of Ro in cell survival after UV irradiation and quality control of small RNAs is not fully understood. Along with the genetic and biochemical characterization of RoRNPs, structural studies in the last decade have significantly increased our understanding of Ro-RNA interaction.

2.1.5 Structural understanding of RoRNP function

Studies on the Y RNAs recognized by Ro in different species have shown that all Y RNAs are predicted to form similar secondary structures (89) with two distinct regions, a conserved domain with almost identical sequences among different species and a variable domain which differs in sequence and predicted structure among species. Based on RNAase cleavage experiments, mutational analysis and secondary structure predictions all Y RNAs are characterized by a helical stem formed by base pairing of the 5' and 3' ends (89). In this stem, there is a highly conserved region, the nucleotide sequence of which is similar in all species. This helix and specifically a bulged cytidine in the helix have been shown to be the Ro binding site (49). Another asymmetric loop in the base paired stem, ~2 bases from the helix has been shown to be essential for Ro

binding. It has been proposed that this loop might be required to widen the helix to allow access of Ro to the binding site. Recently, Christov et al. have shown that Y RNAs are required for chromosomal DNA replication in humans (62). Mutations in the Ro binding site of the YRNAs do not affect DNA replication, suggesting that there are other regions of the Y RNA besides the conserved stem that are involved (62).

Recent structural studies on the eukaryotic ortholog of Ro from *X. laevis* have shown that Ro is composed of two domains (90). One is an elliptical domain formed by numerous repeats of pairs of antiparallel α- helices called HEAT repeats (91). HEAT (named after Huntingtin, Elongation factor 3, A subunit of protein phosphatase 2A and TOR1) repeats are found in a variety of proteins, many involved in cytoplasmic transport processes like GCN1 (transport of tRNA substrates), VP15 (vesicle mediated protein transport) and importins (nuclear-cytoplasmic import and export pathways) (91). While most of these proteins are large in size, the HEAT repeat domain itself could vary from 4 repeats to 36 repeats. *X. laevis* Ro containes 19 HEAT repeats (90). The second domain of Ro resembles the collagen-binding A domain of vonWillebrand Factor (vWFA domain) consisting of a Rossman fold with a β-sheet sandwiched between multiple α-helices (90). The vWFA domain has been well studied in cell-adhesion and extracellular matrix proteins like Integrins, Matrilins and has been identified in many other proteins like the Anthrax toxin receptors and protease regulators etc. (92). In most of these proteins the vWFA domain is involved in mediating protein-protein interactions. Approximately 46% of known vWFA domains also include a divalent cation binding, noncontiguous sequence motif called the MIDAS motif (*M*etal *I*on *D*ependent *A*dhesion *S*ite motif) (93). In prokaryotes, more than 80% of known vWFA domains contain at least an imperfect MIDAS motif (93). It has been shown that coordination of the divalent ion at this site is completed by the interacting protein partner, involving a conserved Aspartate or Glutamate residue (94). In this way the MIDAS motif is thought to act as a 'structural glue'.

The crystal structure of *X. laevis* Ro in complex with a fragment of Y RNA containing the conserved Ro binding site shows that Ro binds Y RNAs at the outer

surface of the helical domain (90). This structure of Y RNA bound Ro from *X. laevis* (90) provides little information about the variable domain of the Y RNA and its interactions with the Ro protein since only the conserved stem of the Y RNA was used in the crystallization experiment. A second RNA binding site has been identified in Ro, at the positively charged central cavity (90), (72). In the structure of RNA bound Ro from *X. laevis,* this site is surrounded largely by the elliptical domain, with the vWFA domain on one end being large enough only to fit a single stranded RNA. It has been shown that this central cavity is the binding site for the single stranded tails of mutant pre-5SrRNAs and U2 snRNAs (90). These studies suggest that Ro possesses two RNA binding sites, one to bind Y RNAs and the other to bind its misfolded RNA substrates. Mutational and structural analysis have also shown that at least part of the mutant pre-5SrRNA binding site on *X. laevis* Ro overlaps the Y RNA binding site, leading to a proposed role for Y RNAs in regulating the binding of Ro to other RNAs like the mutant pre-5SrRNAs or other damaged RNAs (90), (72). The mechanism of Ro binding to misfolded RNA substrates is not yet understood.

In this chapter, we describe the structural studies and biochemical characterization of the prokaryotic ortholog of Ro, namely Rsr from *D. radiodurans*. To obtain structural information to a high resolution, we have used X-ray crystallography. As described in Ramesh A et al., (95) our results show that "while the overall structure of Rsr resembles the architecture of *X. laevis* Ro, there are major structural differences between the two proteins. The most notable difference is at the interface of the two domains, where large movements are evident in the α-helices close to the interface. Movement of these helices results in an enlarged central cavity, suggesting that the structural flexibility of Ro proteins probably aids in switching between 'open' and 'closed' states which may be required for entry, stabilization and release of the RNA substrates at the conserved central cavity. We have also characterized the binding of Rsr to *Deinococcus* Y RNA *in vitro,* using a fluorescence based assay. Our results reveal that Rsr binds Y RNA with at least low-nanomolar affinity and forms a complex with 1:1 stoichiometry. While the apo-Rsr protein is monomeric even at very high

concentrations, electron microscopy and size exclusion chromatography show that upon binding Y RNA Rsr forms multimers of approximately 12 molecules of the 1:1 Rsr:Y RNA complex".

2.2 RESULTS AND DISCUSSION

2.2.1 Sequence based analysis of Ro orthologs

Multiple sequence alignment of Ro proteins from different organisms shows the presence of conserved motifs. At least one conserved RNA recognition motif (RRM) could be identified (**Figure 2.1**). RRMs are a loosely conserved 80 to 90 amino-acid long region, found in proteins that participate in diverse reactions involving RNAs like polyadenylation, splicing, RNA transport and translation (96), (97), (98). The RRM is characterized by the presence of one hexameric and one octameric Ribonucleoprotein (RNP) consensus sequence. In Rsr, these comprise residues 100 to 105 and residues 132 to 139. These RNP consensus motifs are rich in Leucine and Phenylalanine residues which typically stack against the RNA bases.

From SMART (Simple Modular Architecture Research Tool) predictions, including searches within the Pfam database, only one major domain is confidently predicted for Ro. This domain is called the TROVE domain after Telomerase, Ro and Vault (44). This domain is found in TEP1, Ro60 and Vault proteins, that are RNA-binding components of Telomerase, Ro and Vault RNPs respectively and is predicted to be the site of RNA binding (44). In Ro proteins, the RRM lies within the TROVE domain.

Sequence analysis also shows that Ro orthologs contain strictly conserved residues even outside the TROVE domain (from residue 360 to 520).

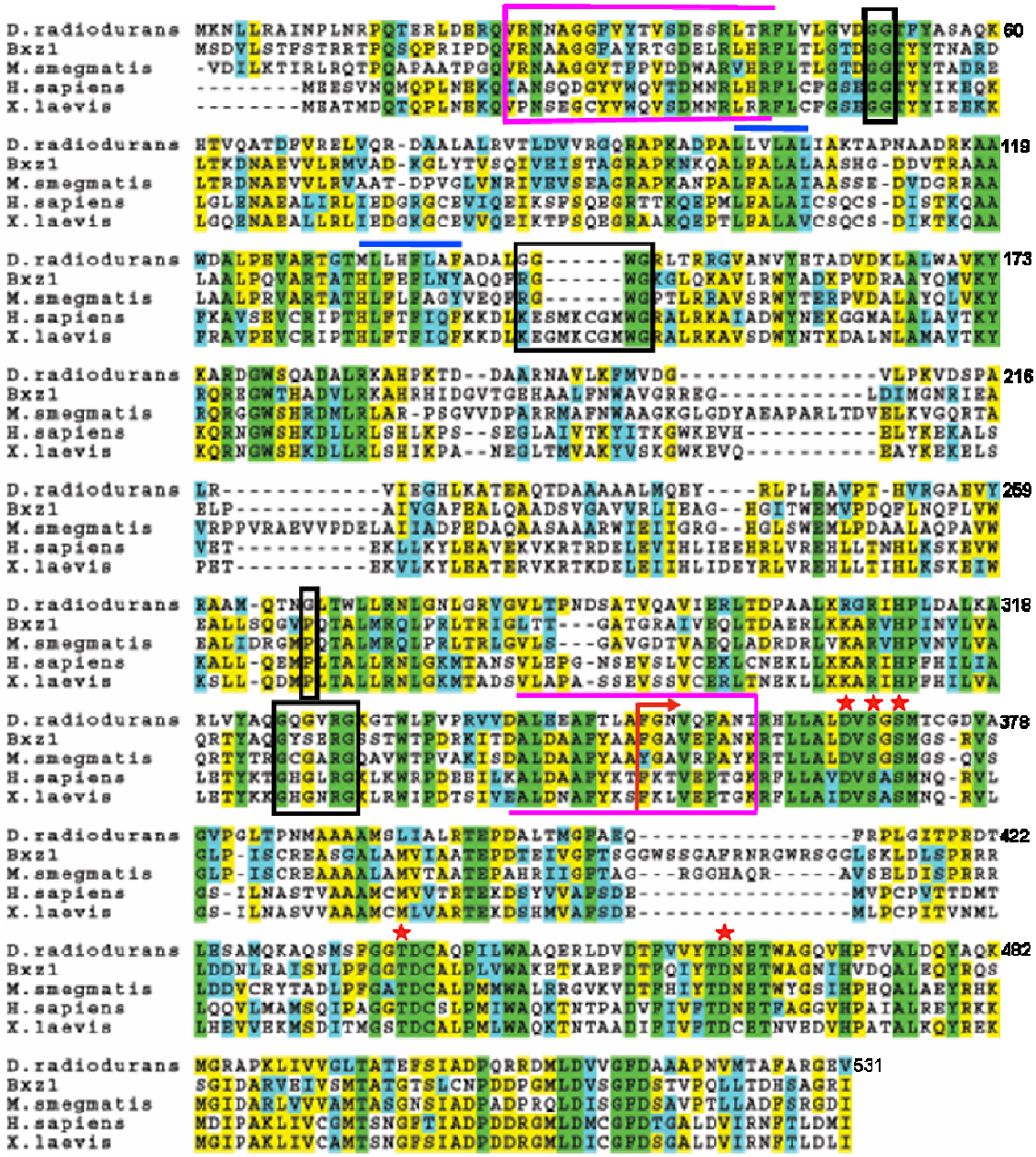


Figure 2.1 Sequence analysis of Ro orthologs

Multiple sequence alignment of Ro proteins using ClustalW with the GONNET weight matrix is shown. *D. radiodurans,* Mycobacteriophage Bxz1, *M. smegmatis*, *X. laevis* and *H. sapiens* Ro proteins are shown with strictly conserved residues in green, identical residues in yellow and similar residues in cyan. Red arrow marks the beginning of the vWFA domain and asterisk marks the MIDAS residues. Magenta brackets mark the ends of the predicted TROVE domain which contain RNP1 and RNP2 (blue lines). Black boxes mark conserved Glycines at flexible hinges in Rsr.

2.2.2 Structure solution of Rsr

The best crystal of Rsr (**Figure 2.2**) diffracted X rays to a resolution of 1.89Å. The first approach was to solve the structure of Rsr using Molecular replacement, with the structure of *X. laevis* Ro as a search model. The two proteins share 37% identity. To aid in the Molecular replacement process, a homology model of Rsr was first created using the web-based program- CPHmodels (99). The coordinates of the homology model were then used for the Molecular replacement. The number of molecules in the asymmetric unit is 1 for both the *X. laevis* and Rsr crystals, hence a single molecule was searched for. Programs like Molrep and Amore and phaser did not find a solution. Considering that the two proteins could differ in structure whereby the two domains of Ro could be oriented differently with respect to each other, the search model was split into two domains and then used for molecular replacement. This also did not yield any clear solutions.

Seleno-methionine substituted Rsr was used for structure solution by Multiwavelength Anomalous Dispersion. The MAD data was complete to 2.6Å and the statistics for data collection are shown in **Table 2.1**. 13 out of 14 sites selenium sites were located by phenix.hyss and the corresponding statistics are shown in **Table 2.2**. Refinement of sites and density modification by AutoSHARP substantially improved the electron density map. This density modified map was used as the input for TEXTAL and the resultant partial model is shown in **Figure 2.3**. Manually cutting and pasting small regions from the CPH homology model created previously, helped to improve the partial model created by TEXTAL. The reasonably complete model was then rigid-body refined against the native 1.89Å data and subsequently refined by simulated annealing starting at a temperature of 5000K, to obtain a higher resolution density map. At various steps, composite omit maps were used to improve the map quality and check for model-bias. 5% of the reflections were kept aside for cross validation of the model through the refinement process. Towards the last stages of model refinement, water molecules were added at a 3σ density cut off, using the corresponding automatic feature in phenix.refine. The added waters were then manually checked and refined. Waters lying on the axis of

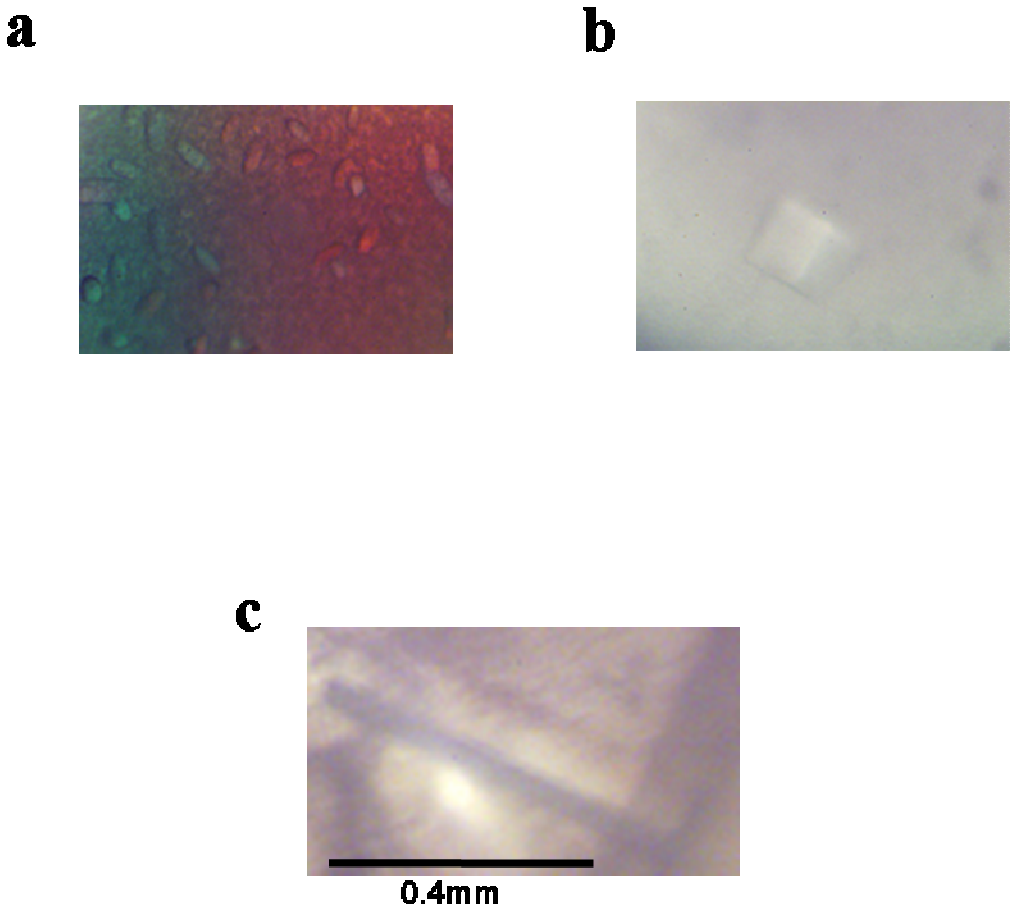


Figure 2.2 Crystals of Rsr

a. Small crystalline pellets formed by 6his-Rsr are shown. b. A single crystal of 6his-Rsr formed in Na/K Tatrate as precipitant is shown. c. A single crystal of cleaved Rsr, grown in PEG 4000 is shown under heavy precipitation of salts.

Table 2.1 Data collection statistics for Rsr and 6his-Rsr:DrY RNA crystals

	Se-Met Rsr			**Rsr**
Space group	*C2*			*C2*
a, *b*, *c* (Å)	105.41, 89.39, 70.55			103.56, 87.61, 70.62
α, β, γ(°)	90.0, 96.82, 90.0			90.0 96.57, 90.0
	Peak	*Inflection*	*Remote*	
Wavelength(Å)	0.97926	0.97939	0.94927	0.97934
Resolution (Å)	50-2.6	50-2.63	50-3.17	50-1.89
R_{sym} or R_{merge}	0.083 (0.64)	0.073 (0.59)	0.106 (0.44)	0.062 (0.69)
$I / \sigma I$	43.4 (1.9)	39.5 (1.2)	24.3 (2.8)	18.9 (1.64)
Completeness(%)	99.6 (96.6)	97.5 (83.2)	96.4 (76.6)	98 (85.3)
Redundancy	7.4 (5.6)	6.7 (3.2)	6.6 (3.5)	3.6 (2.4)

	Rsr:DrY RNA^{Se}
Space group	*P6*
a, *b*, *c* (Å)	187.9, 187.9, 461.7
α, β, γ (°)	90.0, 90.0, 120.0
Wavelength(Å)	0.97935
Resolution (Å)	50- 8.0
R_{sym} or R_{merge}	0.2 (0.76)
$I / \sigma I$	11.22 (1.9)
Completeness (%)	99.7 (99.8)
Redundancy	10.5 (8.3)

Table 2.2 Statistics of the selenium sites in Rsr

Selenium	X	Y	Z	OCCUPANCY	B-FACTOR
Site 1	0.067	0.928	0.537	0.86	60.4
Site 2	0.146	0.904	0.7	0.88	70.1
Site 3	0.179	0.056	0.536	0.78	74.9
Site 4	0.349	0.215	0.061	0.84	91.5
Site 5	0.222	0.16	0.693	0.95	112.1
Site 6	0.29	0.057	0.4	0.73	93.5
Site 7	0.096	0.045	0.78	0.48	72.6
Site 8	0.366	0.346	0.543	0.76	110.4
Site 9	0.376	0.309	0.498	0.46	70.1
Site 10	0.17	0.87	0.64	0.51	83.6
Site 11	0.334	0.989	0.271	0.64	109.6
Site 12	0.376	0.048	0.977	0.49	149.6
Site 13	0.046	0.976	0.314	0.65	176.7

Overall correlation coefficient = 0.691
Overall Z score = 29.66
Overall FOM = 0.66

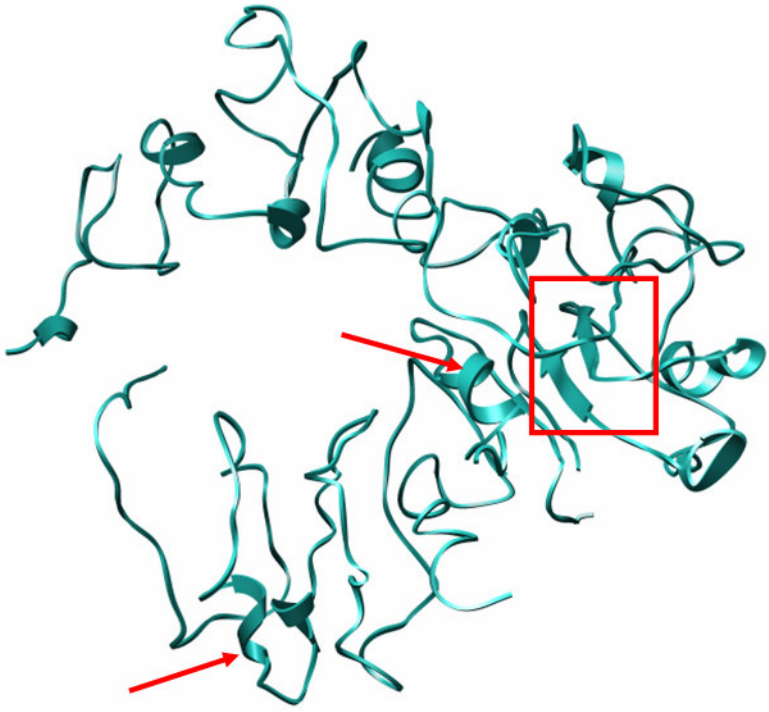


Figure 2.3 Initial model of Rsr built by TEXTAL

Clearly recognizable features like β stands (boxed) and α helices (red arrows) in the initial model are shown.

crystallographic symmetry were refined with "half" occupancy. Strong positive electron density at 6σ contour level was observed, surrounded by residues that are conserved in sequence across the Ro orthologs. This site corresponds to the divalent metal site in *X. laevis* Ro. Since the crystallization solution for Rsr contains 200mM calcium acetate, a calcium ion was fit and refined in this position. The final model refined to an R_{work} of 22.3% and an R_{free} of 26.2% suggesting good correlation between the model and the data. The refinement statistics are shown in **Table 2.3**.

The final model comprises residues 36 to 531 out of the total 531 residues, one Ca^{2+} atom and 400 molecules of water. The model possesses good stereochemistry as judged by the Ramachandran Plot of the backbone φ and ψ angles (**Figure 2.4**). 99% of residues lie in the favoured regions and 1% in the generously allowed regions. Clear electron density was not visible for residues 1 to 35 which are presumably disordered in the crystals. Multiple sequence alignment of Ro from various organisms (**Figure 2.1**) shows that this N-terminal stretch of 35 residues is highly variant in all Ro orthologs except for six conserved residues. In the crystal structure of Ro from *X. laevis,* this region is placed on the outer surface of the protein, away from the RNA binding sites and forms two short β-strands which pack against a β-strand of the C-terminal domain (90). The crystallographic asymmetric unit consists of one monomer of Rsr and applying crystallographic symmetry does not generate any higher oligomer of Rsr with significant buried surface area. This, along with size-exclusion studies performed on purified Rsr suggests that Rsr is a monomer.

2.2.3 The overall structure of *D. radiodurans* Rsr

The overall structure of Rsr resembles a ring spanning ~ 76 Å in diameter and ~ 32 Å in thickness (**Figure 2.5**). Rsr possesses two distinct domains: an N-terminal domain including residues 36 to 351 and a C-terminal domain including residues 352 to 531. The N-terminal domain comprises of numerous short helices (H1 to H18) which form a ringlike arrangement, and one beta-strand (β1). The ringlike arrangement of helices encloses a central cavity with dimensions of 18 to 20 Å and is lined by charged residues

Table 2.3 Refinement statistics for the Rsr structure

Resolution (Å)	46.6-1.89
No. reflections	48610
R_{work} / R_{free}	22.3 / 26.2 (42.2 / 49.7)
No. atoms	
Protein	4141
Ca	1
Water	400
***B*-factors**	
Protein	25.3
Ca	5.1
Water	28.9
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.71

Highest resolution shell is shown in parenthesis

$$R = \frac{\Sigma_n \| F_{obs} - F_{calc} \|}{\Sigma_n |F_{obs}|}$$

Where n represents the set of reflections used for refinement.

5% of reflections were used for the calculation of R_{free}.

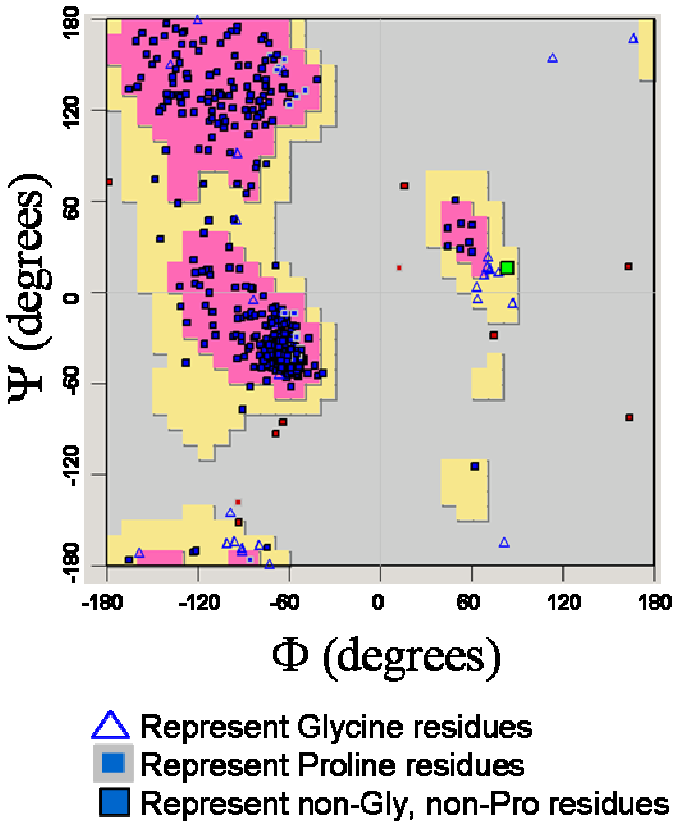


Figure 2.4 Ramachandran Plot of the Rsr crystal structure

Backbone torsion angles of the Rsr structure are shown.

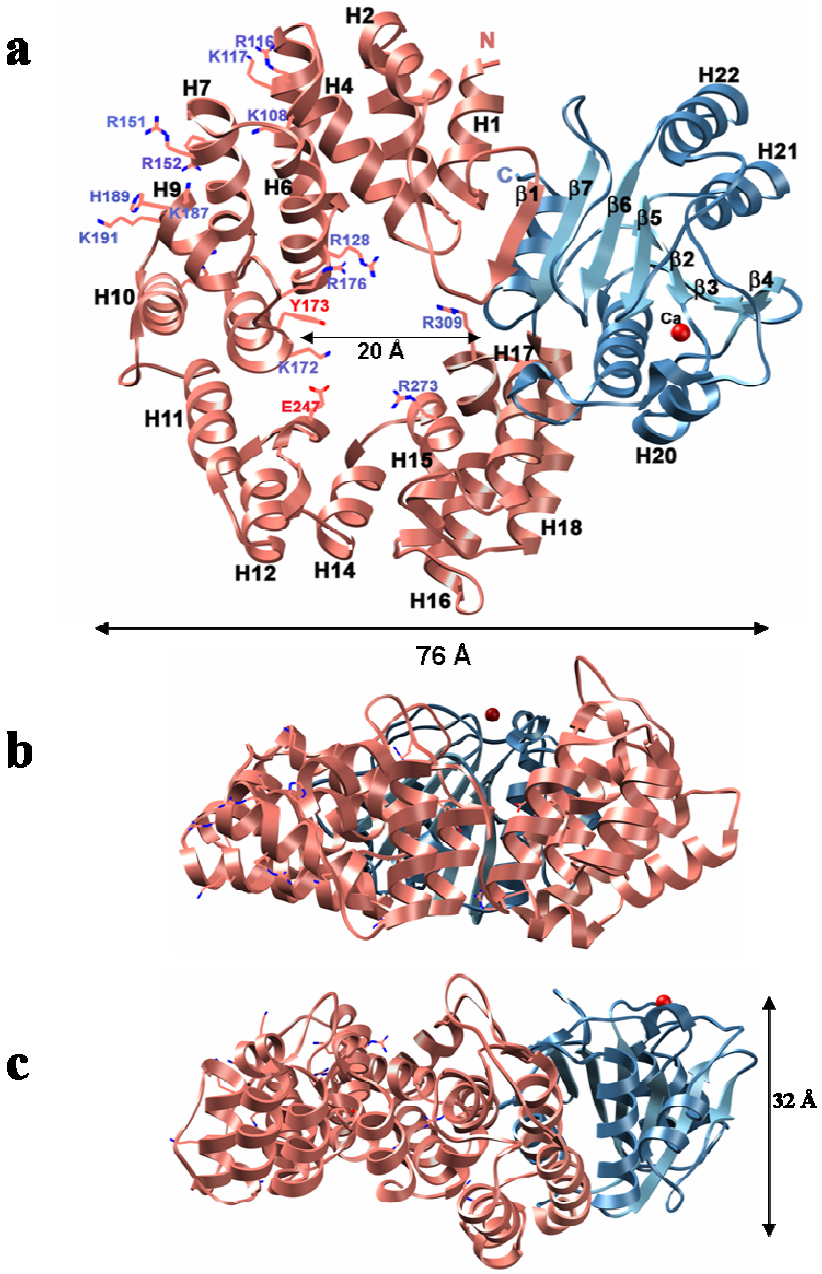


Figure 2.5 Overall structure of *D. radiodurans* Rsr

a. The N-terminal ring domain (pink) consists of helices H1 to H18 and one beta strand- β1. The C-terminal vWFA domain (blue) consists of a β2 to β7, sandwiched between helices H19 to H23. The Calcium ion is shown as a red sphere. Conserved residues in the central cavity (Arg128, Lys172, Tyr173, Arg176, Glu247, Arg273 and Arg309) and on the outer surface of the ring domain (Arg116, Lys117, Arg151, Arg152, Lys187, His189, Lys191) are shown as sticks. b. Side view of Rsr when looking from H10-H11. c. Side view of Rsr when looking from H12-H16.

from the loops connecting successive helices from H2 to H18. The majority of residues pointing into this cavity, for example - Arg128, Lys172, Tyr173, Arg176, Glu247, Arg273 and Arg309 are strictly conserved among the Ro orthologs. In Ro from *X. laevis*, this site has been implicated in binding single stranded RNAs with these conserved residues directly interacting with an extended conformation of a single strand of RNA (90). Notably, on the outer surface of the helical ring domain there is another cluster of positively charged residues (Arg116, Arg117, Arg151, Arg152, Lys187, His189, Lys191), some of which are highly conserved in Ro orthologs. In *X. laevis* Ro, this site has been implicated in Y RNA binding (90).

The short repeats of antiparallel helices as seen in the N-terminal domain of Rsr are similar to the HEAT repeats seen in the nuclear transport proteins like the importin family of proteins (91), (100), (101). **Figure 2.6** shows the arrangement of HEAT repeats in other proteins. The structural similarity between these HEAT repeat domains and the N-terminal region of Rsr suggests a similar functional mechanism for the domain in these proteins.

The C-terminal domain of Rsr resembles a vonWillebrand Factor A (vWFA) domain similar to that found in cell-adhesion and extra-cellular matrix proteins (93). The vWFA domain is a Rossman fold, characterized by a β sheet sandwiched between α-helices. In Rsr, the β-sheet is formed by six parallel β-strands (β2 to β7) and is sandwiched by parallel helices H19, H20 and H23 on one side and helices H21, H22 on the other (**Figure 2.5**). Interestingly, in Rsr this C-terminal β –sheet interacts with the β-strand (β1) from the N-terminal domain, to form a seven stranded sheet. The vWFA domain is found in a variety of other proteins, where it assumes a common fold (shown in **Figure 2.7**). Frequently, the vWFA domain is the site of protein-protein interaction, for example, in the vonWillebrand Factor this domain helps the adhesion of platelets to fibrillar collagen underlying injured vascular endothelium and in the Inserted (I) domain of Integrins, this domain is involved in interactions with collagen (92).

About 46% of known vWFA domains also contain a non-contiguous sequence motif called the MIDAS motif, which forms a divalent metal binding site (102).

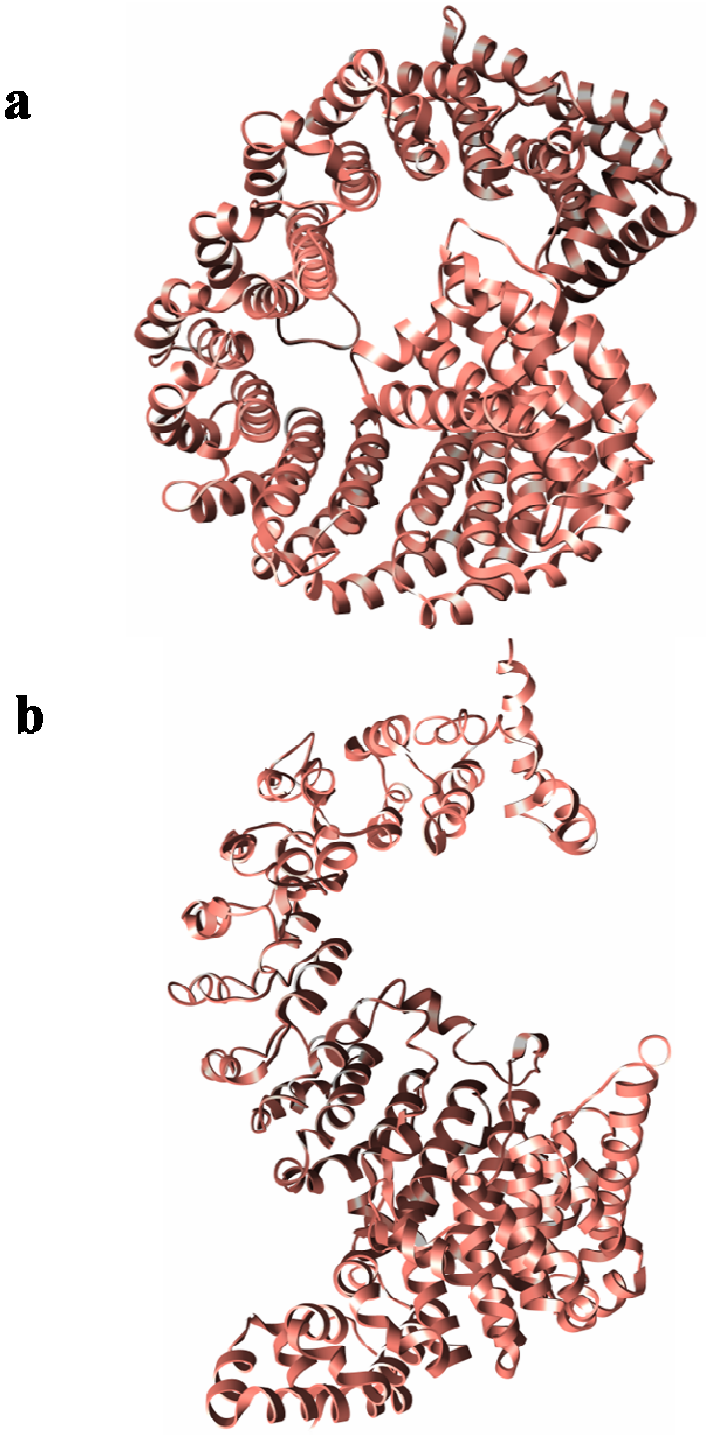


Figure 2.6 HEAT repeats in nuclear transport proteins

a. HEAT repeats (salmon) in crystal structure of importin-β bound to IBB domain of importin-α (not shown), 1QGK. b. HEAT repeats (salmon) in the crystal structure of karyopherin β2-transportin in complex with Ran (not shown), IQBK.

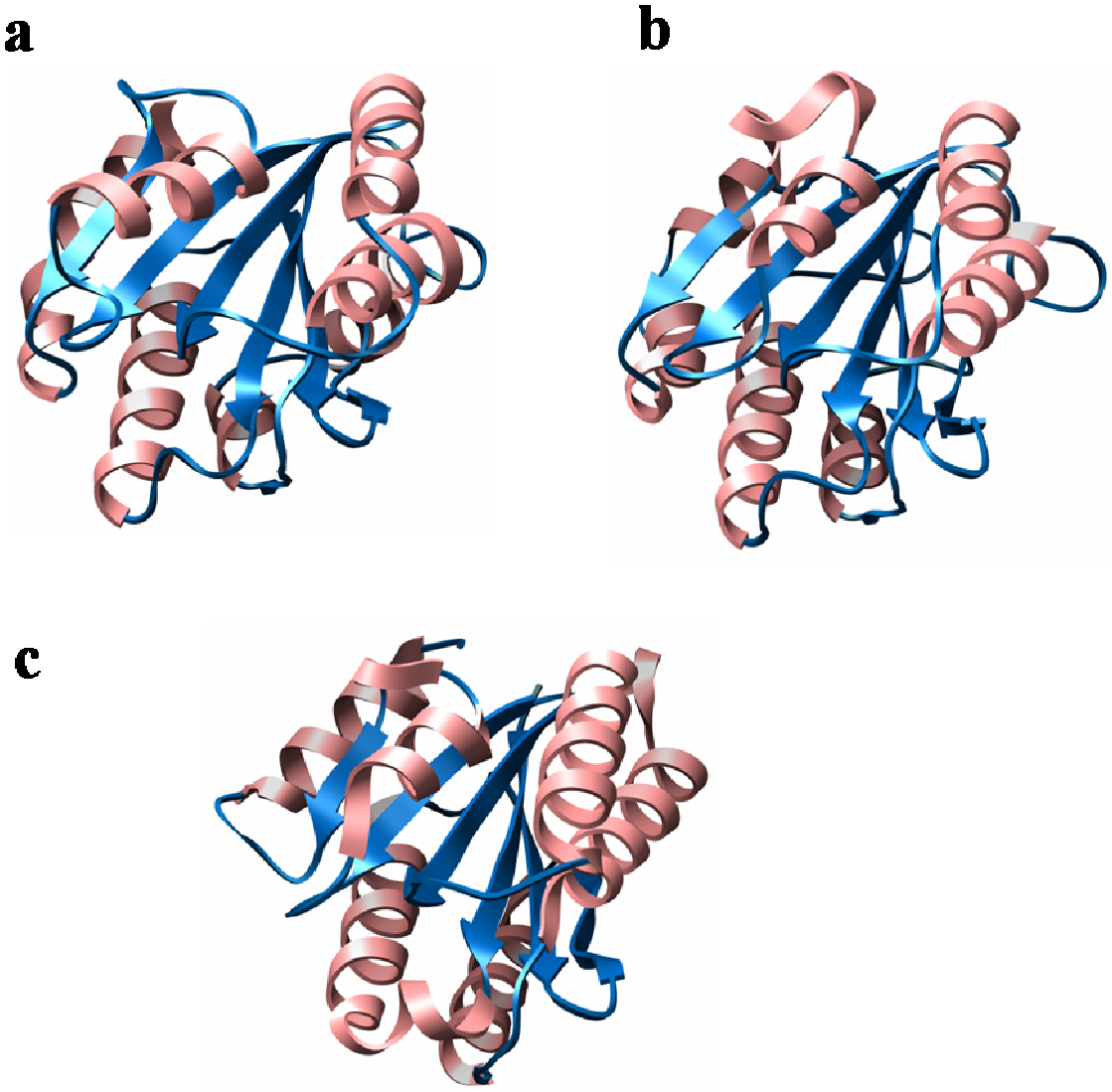


Figure 2.7 vWFA domains in proteins

a. vWFA domain in the crystal structure of the I domain of integrin bound to Mn^{2+} (1JLM). A six stranded parallel β-sheet (blue) sandwiched between α-helices (salmon). b. Crystal structure of the A3 domain of the von Willebrand factor, 1AO3 shows a 6-stranded β-sheet (blue) between α-helices (salmon). c. vWFA domain in the crystal structure of the human capillary morphogenesis protein 2, an anthrax toxin receptor, 1SHT. A six stranded β-sheet (blue) is sandwiched between α-helices (salmon).

Similarly, the C-terminal domain in Rsr also includes the divalent cation binding, MIDAS (*M*etal *I*on *D*ependent *A*dhesion *S*ite) motif. This motif consists of residues $D^{367}XSXS\ldots T\ldots D^{462}$, all of which are strictly conserved in all Ro orthologs. In Rsr, a peak of strong positive electron density corresponding to the 6σ contour level in a difference density map was proximal to a site surrounded by the MIDAS residues. A Ca^{2+} ion was fit into this site based on the fact that the crystallization condition contains 200mM Calcium acetate (**Figure 2.8**). The Ca^{2+} ion is coordinated by O^{γ} of Ser369, O^{γ} of Ser371, two structured water molecules (HOH 3 and HOH 6), $O^{\varepsilon 2}$ of Glu464 and $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$ of Glu464 from a crystallographically symmetry related molecule. The distances between the Ca^{2+} ion and the ligands range from 2.3 to 2.6Å. Two residues of the MIDAS motif- Asp367 and Asp462 form the second coordination sphere around the Calcium ion and bind the metal coordinating waters. Several crystal structures are available, of vWFA domains containing the MIDAS motif. This feature of a metal coordinating ligand belonging to a symmetry related molecule has been observed in other crystal forms of the MIDAS motifs in vWFA domains and Integrin I domains (103), (104) where it has been proposed that in physiological conditions the coordinating ligand may belong to a binding partner of the protein. While a disulphide bond between Cys374 of symmetry related molecules of Rsr was observed, this is likely to be an artifact of crystallization where a high concentration of protein was used. In the RNA bound structure of *X. laevis* Ro, the metal site is occupied by a Mg^{2+} ion and the metal coordination is completed by an acetate molecule from the crystallization solution. It has been suggested that the acetate may mimic an aspartate or glutamate residue from a yet unidentified binding partner of Ro (90). In the apo structure of *X. laevis Ro,* this domain is structurally identical to the metal bound *X. laevis* and Rsr structures (72). The authors have refined a water molecule in place of the metal.

In both Integrins and the cell-adhesion proteins, the vWFA domain is the site for protein-protein interactions with a conserved cation binding MIDAS site acting as a structural 'glue', bringing protein ligands closer to the vWFA domain (94). The coordination of the metal is made complete by ligands from the interacting partner. In

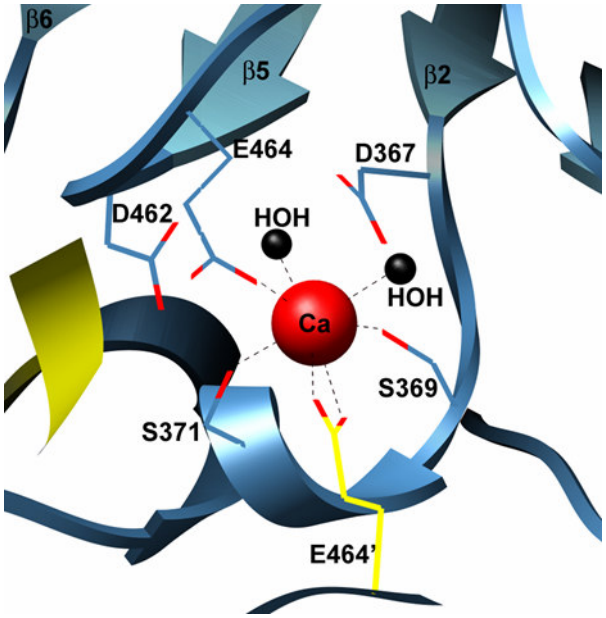


Figure 2.8 Close-up of the divalent metal site

Coordination sphere around Calcium (red sphere) bound to Rsr is shown. The Calcium coordinating residues – Ser 369, Ser 371, two waters, Glu464 (sticks) along with Glu464' from the symmetry related molecule (yellow stick) are shown. The second coordination sphere around Calcium, consisting of MIDAS residues Asp 367 and Asp 462 which stabilize the water molecules are shown as sticks.

many cases, metal binding at the MIDAS site is also associated with large conformational changes at a distant site across the domain (105). Based on the similarity in the fold of the C-terminal domain of Rsr to other vWFA domains, it is possible that this conserved domain of Rsr along with the metal site helps to mediate interactions between Rsr and other proteins. Also, since Rsr has been shown to bind small RNAs that are damaged or targeted for degradation (83), it is possible that binding partners of Rsr include nucleases, helicases and other RNA processing enzymes. Also, isolation of Ro RNPs from eukaryotic cells often pulls down other proteins like the La protein and the 52kDa Ro protein. If at all Ro interacts directly with these proteins, it is possible that the vWFA domain mediates these interactions.

2.2.4 Rsr binds Y RNA with low nanomolar affinity

To characterize the binding of Rsr to *D. radiodurans* Y RNA, a fluorescence based assay was used. Rsr has eight tryptophans, five of which are strictly conserved in all Ro homologs (**Figure 2.9**). Also, the structure of *X. laevis* Ro bound to RNA (90) suggests that one or more tryptophans are involved in stacking interactions with RNA. Besides the full length DrY RNA, truncations of DrY RNA were used in this experiment (**Figure 2.10**). These included 35- mer and 33-mer fragments representing the putative Rsr binding site of DrY RNA (DrY RNA^{35}), and a 39-mer fragment with an engineered, stable GAAA tetraloop (called DrY RNA^{GAAA}). By monitoring the quenching of intrinsic tryptophan fluorescence upon titrating RNA, binding isotherms were obtained (**Figure 2.9**). The binding assay was performed in buffer containing 25mM Hepes pH 7.0, 150mM sodium chloride and 10mM magnesium chloride. The minimum detectable amount of Rsr was found to be 25nM, hence the assay was done at 50nM concentration of Rsr. Relevant corrections for RNA fluorescence, photobleaching and inner-filter effect of RNA were performed as described in **section 2.3.12**. Using this assay it was found that Rsr binds DrY RNA^{35} and DrY RNA^{GAAA} with at least 16±3nM and 36nM affinity respectively with a stoichiometry of 1:1. The affinity for the full length DrY RNA is lower, with a K_d = 280nM. This could be the effect of the variable domain of

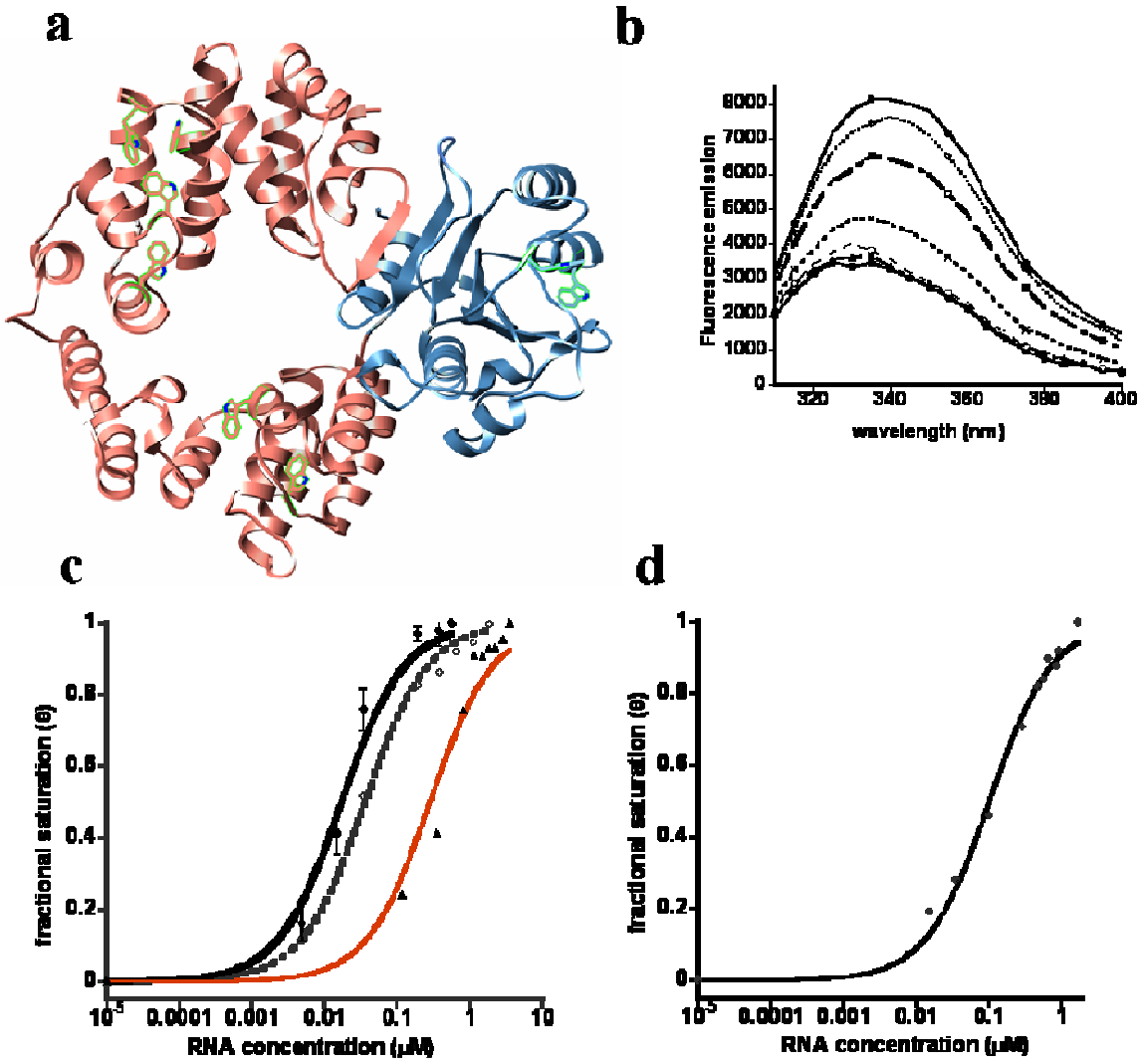


Figure 2.9 Binding studies on Rsr and Y RNA

a. Tryptophan residues are shown on the Rsr structure, as sticks. b. Fluorescence emission spectra (315 to 400nm) obtained upon excitation at 295nm are shown. Quenching of fluorescence occurs with increasing concentrations of Y RNA. c. Plot of fractional saturation (θ) vs. Y RNA concentration (μM) obtained from the emission at 335nm. Binding with DrYRNA^{35} (black), DrY RNA^{123} (red), DrY RNA^{GAAA} (dotted black) are shown. Error bars represent standard deviation from the mean of three independent samples. The equilibrium dissociation constant Kd equals ~ 16 ± 3 nM for DrY RNA^{35}, 36nM for DrY RNA^{GAAA} and 280nM for Dr YRNA^{123}. Stoichiometry of binding is 1:1. d. Binding isotherm for Rsr TROVE domain with Dr YRNA^{35} is shown. K_d = 100nM.

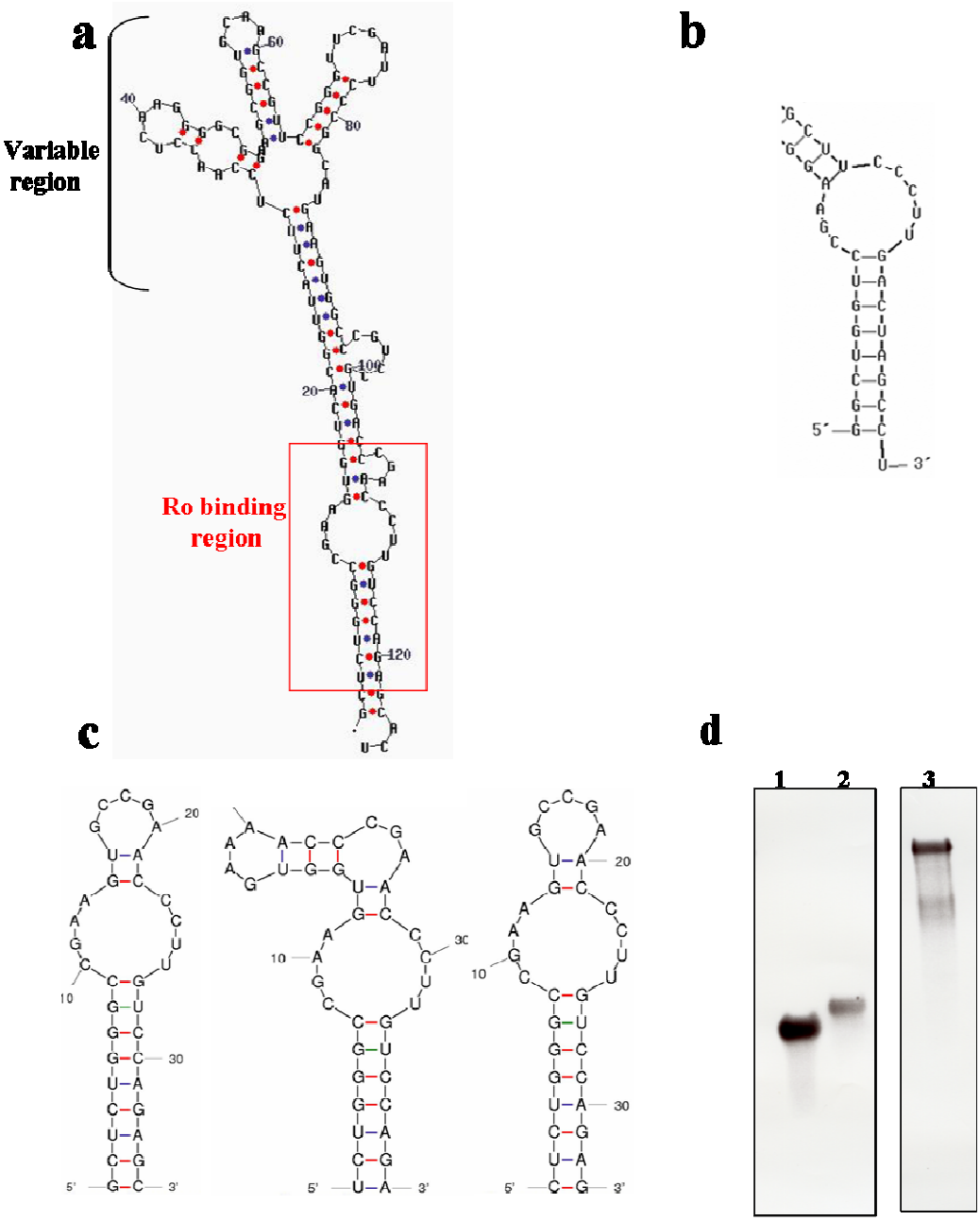


Figure 2.10 Potential secondary structures of Y RNAs as predicted by mFOLD

a. The structure of full length DrY RNA^{123} is shown and the variable domain and conserved Ro binding regions are marked. b. Ro binding region of *X.laevis* Y RNA is shown. c. mFOLD secondary structure predictions of the 35-mer (DrY RNA^{35}), 39-mer (DrY RNA^{GAAA}) and 33-mer (DrY RNA^{33}). d. Native PAGE analysis of DrY RNA^{35} (lane1), DrY RNA^{GAAA} (lane2) and DrY RNA^{123} (lane3) shows presence of predominantly one conformation.

the Y RNA which is absent in the tighter-binding truncations. It is possible that the conformational flexibility of the variable domain may affect the binding by Rsr whereas in physiological conditions this variable domain may be associated with other cellular components. Since this assay could not be performed at a lower concentration of Rsr because of the detection limit, the calculated K_d is likely to represent the apparent K_d. RNA binding affinity was weaker in the presence of 250mM lithium sulfate instead of 150mM sodium chloride, suggesting a salt-dependence of Ro-RNA binding. Consistent with our findings, dissociation of RNAs from Ro at higher ionic strength has also been shown for human Ro ribonucleoprotein particles (106). Binding of Rsr to DrY RNA^{35} is comparable to that reported for *X. laevis* Ro and Y RNA interactions. *X. laevis* Ro binds Y RNA and the Ro binding region with ~5nM affinity. In *X. laevis* it as been shown that mutation of the conserved Histidine residue (H189 in Rsr and H187 in *X. laevis* Ro) to Serine, on the outer positively charged surface of Ro, decreases the affinity for Y RNA. In Rsr, this is being tested.

To study the binding of Y RNA by the TROVE domain alone, in the absence of the vWFA domain, TROVE was purified from inclusion bodies as described in the methods section. After removal of Empigen BB from the protein sample, fluorescence assays were performed. **Figure 2.9** shows the binding of the TROVE domain with DrY RNA^{35} with an apparent K_d = ~ 100nM. Although the TROVE domain alone binds Y RNA, the affinity is almost 6-fold weaker than that of full length Rsr. This suggests that the vWFA domain may be involved in stabilizing the Y RNA binding conformation of the TROVE domain. As seen in the crystal structure, the TROVE domain consists of multiple HEAT repeats and is expected to be conformationally flexible (107). The presence of the vWFA domain may aid in closing-in of the HEAT repeats to stabilize the ringlike conformational of the TROVE domain that may be suitable for RNA binding.

2.2.5 Rsr shows large domain movements when compared with apo or RNA bound Ro from *X. laevis*

Consistent with our results that molecular replacement using *X. laevis* Ro (1YVP and

1YVR) with the diffraction data from Rsr was unsuccessful, the comparison showed very obvious structural differences between the two proteins. The overall shape of the two proteins is significantly different, with Rsr forming a more rounded ring shaped N-terminal domain compared to the elliptical N-terminal domain of *X. laevis* Ro (**Figure 2.11**). The side views of the two proteins clearly show the flat ring formed by Rsr compared to the V-shaped, distorted ring formed by the *X. laevis* Ro. The surface rendering of Rsr when colored by electrostatic potential shows two distinct patches of positive residues, one on the outer face of the ring domain corresponding to the Y RNA binding site of *X. laevis* Ro and the other lining the central cavity which corresponds to the single stranded RNA binding region in *X. laevis* Ro. While the residues in these 2 sites are conserved, in *X. laevis* Ro the distribution of the positive surface is more continuous between the 2 sites, unlike the separate clusters seen in Rsr. The difference in charge distribution could account for the variations in RNA sequences that bind Rsr versus *X. laevis* Ro. In both proteins, the vWFA domain is mostly negatively charged.

Superimposition of the entire proteins as rigid bodies, using COOT (108) demonstrated that the relative position of the two domains with respect to one another was quite different, with RMS deviations based on Cα positions equal to 3.8 Å. The superimposition using the Y RNA binding region as reference, is shown in **Figure 2.12**. A much better fit was obtained when the domains were treated as two rigid bodies and superimposed on the corresponding domains from *X. laevis* Ro with RMS differences of 2.8 Å for the N-terminal and 2.1 Å for the C-terminal domains.

Superimposition of the vWFA domains of Rsr and *X. laevis* Ro (**Figure 2.13**) shows no pronounced changes within the domain. One difference is in the loop region between residues Gln373 to Pro385. This loop is close to the metal binding site and the changes seen could be a result of the presence of the larger Ca^{2+} ion in the metal site as compared to a Mg^{2+} ion.

However, superimposition of the N-terminal domains showed marked differences between the structures at the interface of the N and C terminal domains (**Figure 2.12 and Figure 2.14**). In Rsr, the stretch of residues from Gly52 to Ala58 that forms a β-

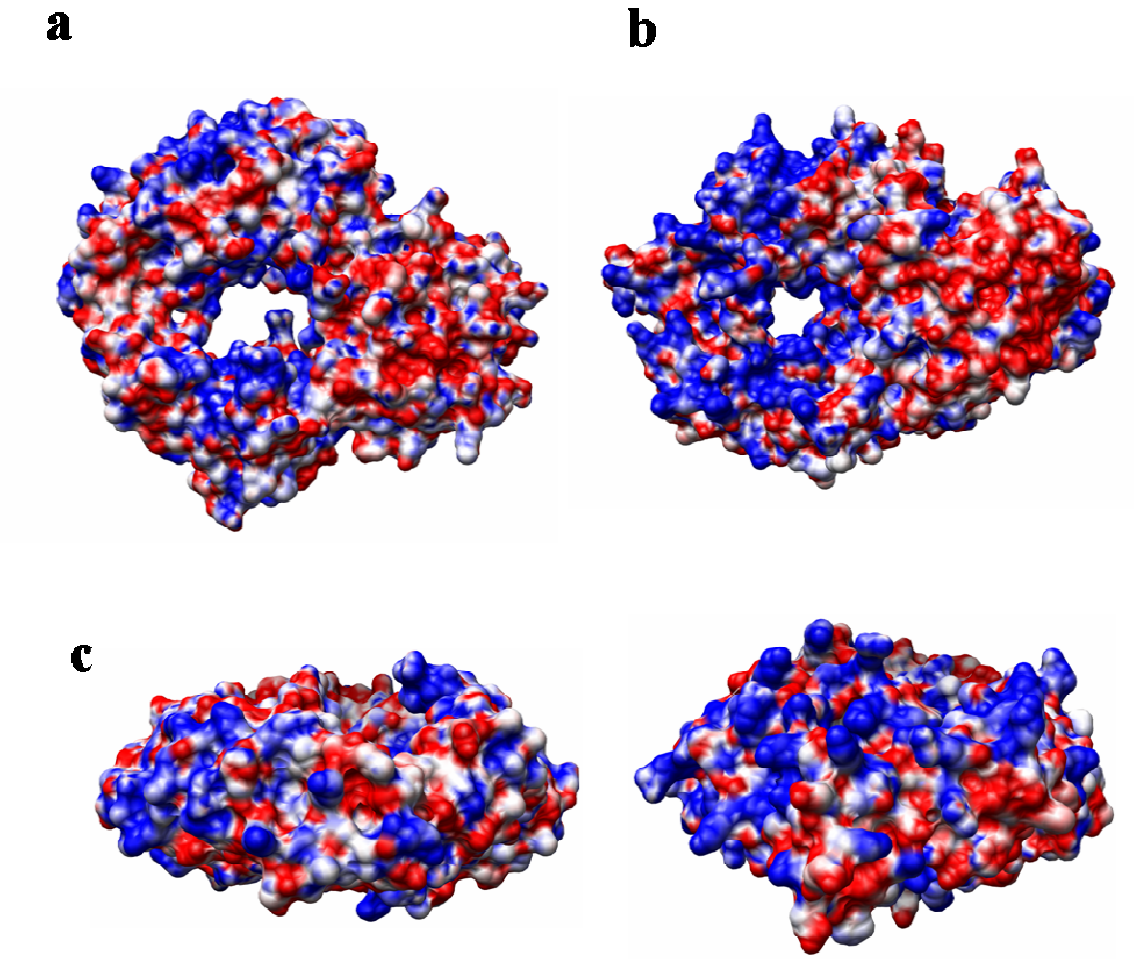


Figure 2.11 Comparison of Rsr and *X. laevis* Ro surfaces

a. The electrostatic surface potential rendering of Rsr shows two distinct positively charged surfaces (blue) corresponding to the residues shown as sticks in Figure 2A. The vWFA like domain is mostly negatively charged (red). b. The surface of *X. laevis* Ro shows a continuous patch of positive charge on the ring domain and a vWFA domain that is mostly negatively charged. c. Side views of Rsr (left panel) and *X. laevis* Ro (right panel) show obvious difference in shape and thickness, with Rsr forming a flat disc with an accessible surface area of 23570 Å^2 as compared to the cup shaped *X. laevis* Ro with an accessible surface area of 23297 Å^2.

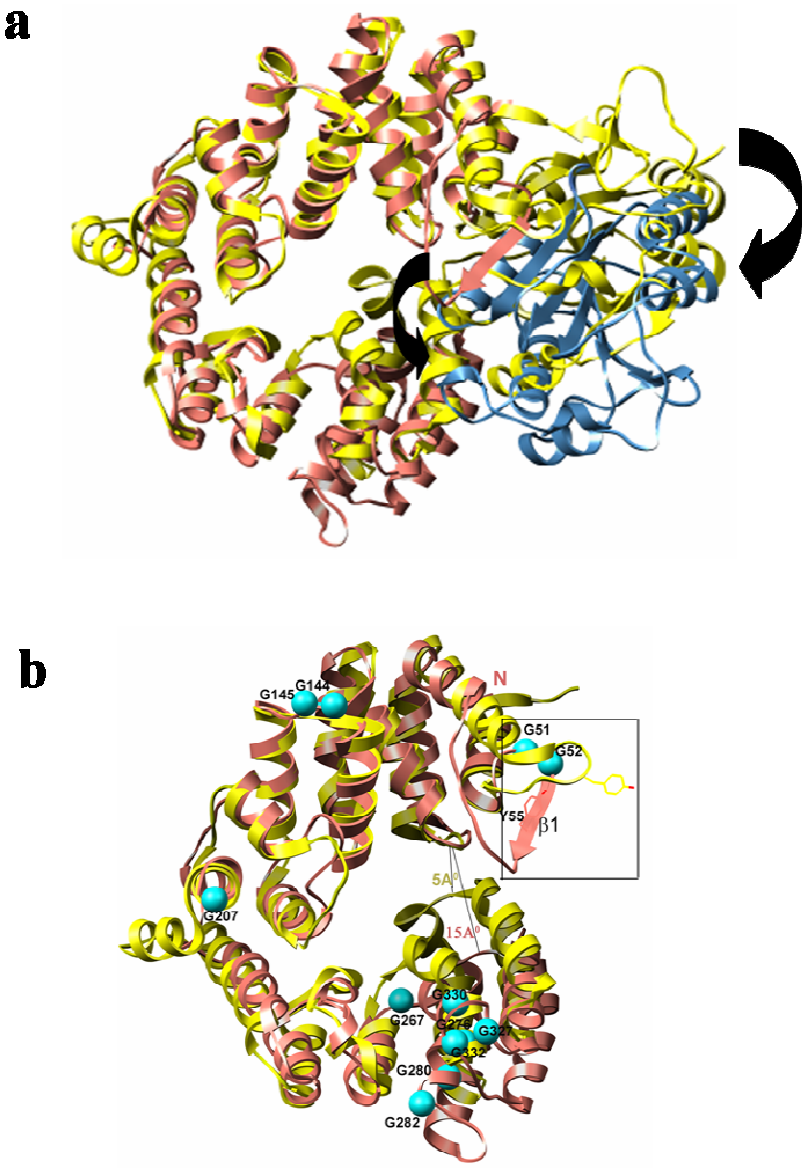


Figure 2.12 Superimpositon of Rsr and *X. laevis* Ro structures with the Y RNA binding region as reference

a. Superimposition of Rsr (pink and blue) on *X. laevis* Ro (yellow) shows large movements of the vWFA domain and helices close to the central cavity. Black arrows mark the direction of movement. b. Close-up of superimposed helical domains of Rsr (pink) and *X. laevis* Ro (yellow) shows H15 to H18 in Rsr are 15Å away from H3 - H4 (black line) whereas in *X. laevis* Ro this distance is ~5Å. The β1 region of Rsr forms an alternate loop conformation in *X. laevis* Ro (boxed region). The conserved tyrosines (Tyr55 in Rsr , Tyr47 in *X. laevis* Ro) have been shown as sticks. Conserved glycine residues have been shown as cyan spheres.

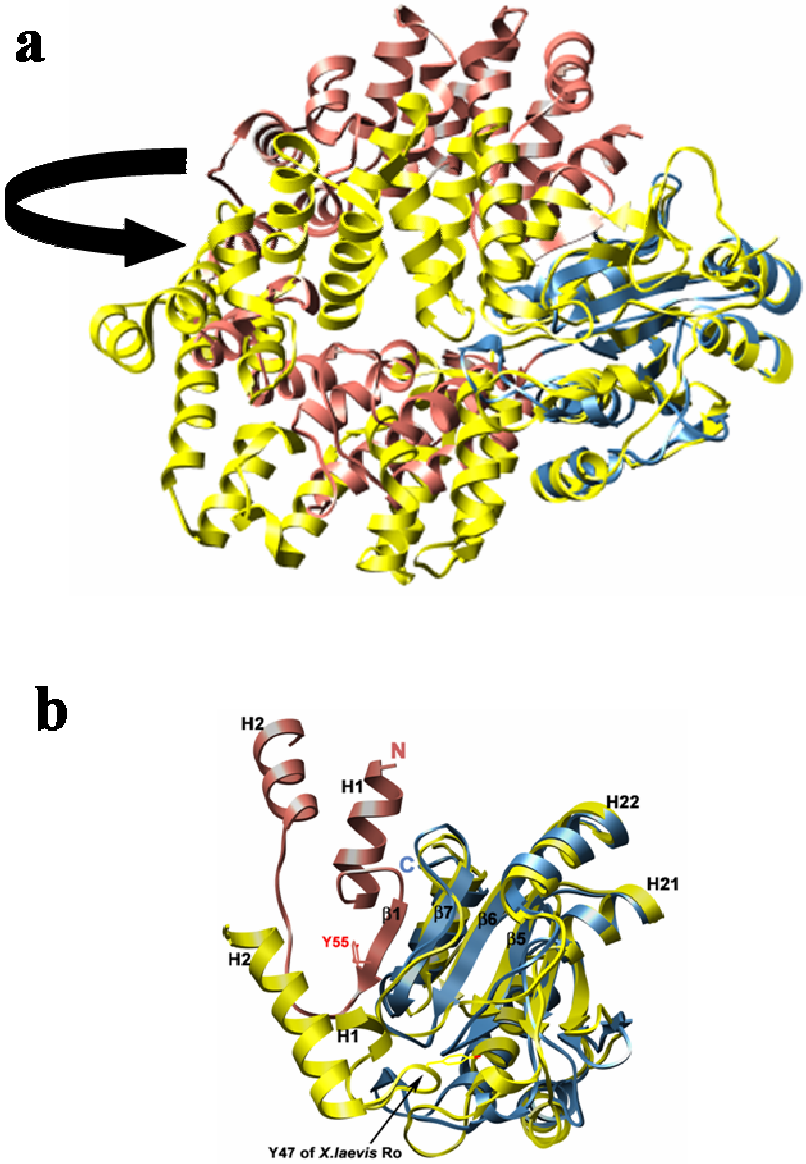


Figure 2.13 Superimposition of Rsr on *X. laevis* Ro with vWFA domains as reference

a. Superimposition of Rsr (pink and blue) and *X. laevis* Ro (yellow) shows displacement of the ring domain, marked by the black arrow. b. Close-up of the superimposed C-terminal domains of Rsr (blue) and *X. laevis* Ro (yellow) have been shown along with the β1 region of Rsr and the corresponding loop from *X. laevis* Ro. In Rsr Tyr55 is placed away from the metal site and is stabilized by interactions with a water molecule and Glu498 whereas in *X. laevis* Ro the corresponding Tyr47 interacts with a metal coordinating water molecule.

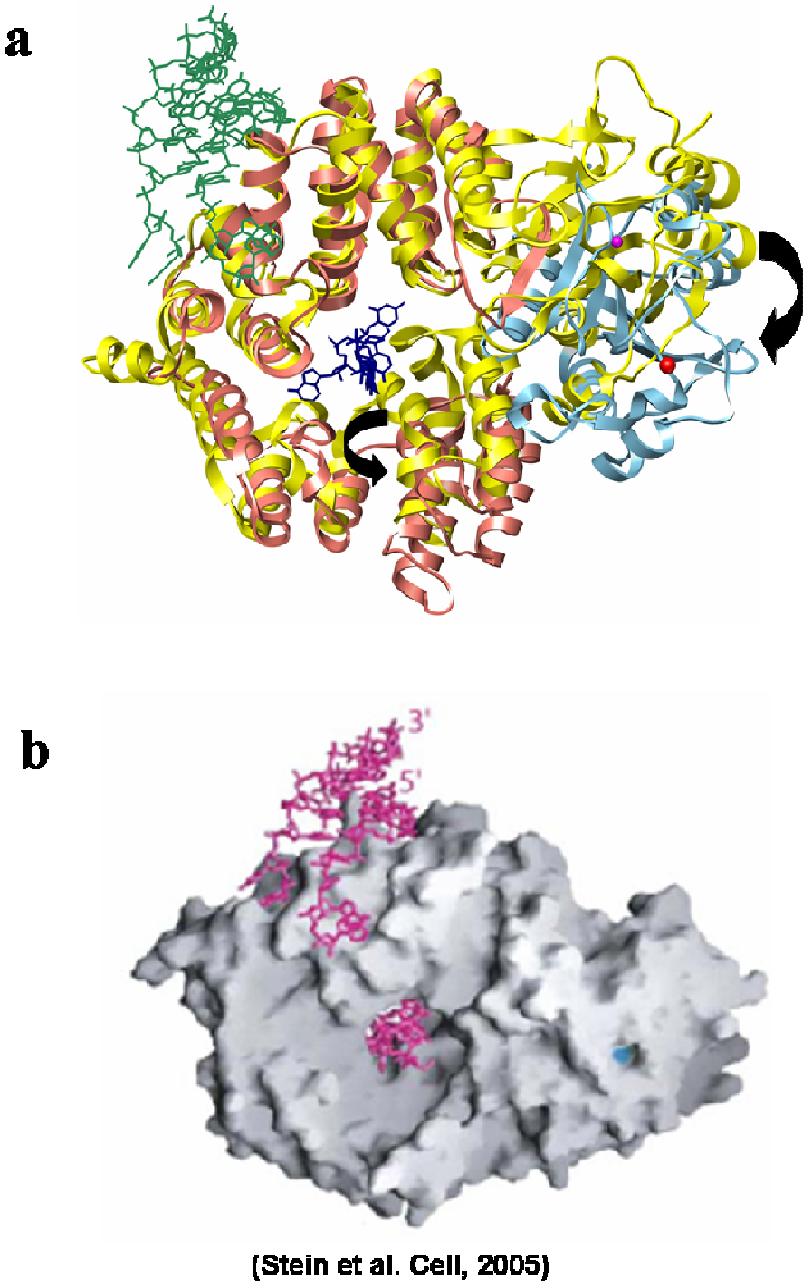


Figure 2.14 Comparison of Rsr and *X. laevis* Ro structures with respect to the RNA bound
A. Superimposition of Rsr (pink and blue) and RNA bound *X. laevis* Ro (yellow) using helical domains as reference shows displacement of vWFA domain and helices H15 to H18. Direction of displacement is shown as black arrows. Double stranded Y RNA fragment (green), ssRNA (navy) and Mg^{2+} ion (magenta) bound to *X. laevis* structure are shown. B. RNA (pink stick representation) bound to *X. laevis* Ro (surface representation) is shown.

strand (β1) is stabilized as a part of the seven stranded β-sheet from the vWFA domain whereas the corresponding region in *X. laevis* Ro forms a loop. In *X. laevis* Ro, a conserved tyrosine residue (corresponding to Tyr55 in Rsr) in this loop has been shown to form the second coordination sphere around the MIDAS metal, forming a hydrogen-bond with one of the metal- coordinating water molecules. However, in Rsr this region is far from the metal site with Tyr55 pointing away from the β-sheet and forming hydrogen bonds with a water molecule and Glu498 (**Figure 2.12**).

Comparison of the N-terminal domains also shows a much larger central cavity in Rsr with the 4-helix bundle formed by H15 to H18 being placed > 15Å away from helices H3 - H4 as compared to the structure of *X. laevis* Ro where this distance is ~ 5Å (**Figure 2.12**). This 4-helix bundle undergoes >34º rotation away from the central cavity as compared to *X. laevis* Ro. The different placement of these helices in Rsr could potentially modulate RNA binding by influencing the size of the central cavity. Considering that the α-helices H17 - H18 are in close contact with the C-terminal domain, changes in the placement of these helices may get transmitted to the C-terminal vWFA domain. It is possible that this structure of Rsr with the 4-helix bundle formed by H15 to H18 placed away from the central cavity, may represent an "open" state poised to allow entry or release of the RNA substrates from the central cavity. In the presence of the RNA substrate as seen in the *X. laevis* Ro bound to the ssRNA (90) the 4-helix bundle may swing in towards the cavity, accompanied by the movement of the vWFA domain which may further affect the down-stream events of RNA processing.

To understand the basis of such flexibility in the structure of Rsr, a detailed analysis of its sequence was done. Rsr has a distribution of highly conserved Glycine residues close to the regions that involve structural changes (**Figure 2.12**). One example is the Gly51-Gly52 motif preceeding the β-strand (β1) which is present in an alternate loop conformation in *X. laevis* Ro. Another Gly144-Gly145 motif precedes the region connecting helices H6 and H7. This region corresponds to a loop in *X. laevis* Ro that becomes ordered upon Y RNA binding (90). Especially notable are Gly267 (conserved Proline or Glycine at this position) which precedes the 4-helix bundle (H15 to H18), and

the Gly-X-X-Gly-X-X-Gly (276-279-282) motif that is present on the C-terminus of helix H15. Glycine-rich motifs in other proteins have been shown to be the site of structural flexibility (109),(110). The positions of these Glycines in Rsr possibly represent hinge regions which allow helices H15 to H18 to undergo displacement away from the central cavity. The N-terminal domain of Rsr is composed of HEAT repeats very similar to that found in some nuclear-transport proteins like importin-β. Various crystallographic, biochemical and molecular dynamics simulation studies strongly support a "spring-loaded" model for ligand binding which involves the movement of some helices with respect to the others with a critical role for glycine and proline residues as hinge residues (107). Our proposal that the structure of Rsr represents the "open" form and the conformational differences we see at the RNA binding central cavity agrees well with this well-described mechanism for ligand binding by HEAT repeat domains in nuclear transport proteins. To confirm this prediction, extensive mutational analysis of these hinge glycine residues is required, along with reliable assays that could help measure the flexibility of specific regions of the protein in solution. Fluorescence based assays to measure conformational changes are sensitive, but the ability to introduce probes into desired locations within the protein is crucial for the outcome of these experiments and often the most challenging part. In Rsr, the presence of multiple reactive groups along with the presence of an extensive and buried interface adds yet another level of complexity.

2.2.6 Potential implications of the Rsr structure on its interactions with different RNA substrates

To understand the features of Rsr that allow its interaction with different RNA substrates, we compared the structure of Rsr with the *X. laevis* Ro bound to Y RNA and single-stranded (ss) RNA (1YVP) (**Figure 2.14**) . At the Y RNA binding site, residues that interact with the nucleotides G1 to A10 in the first strand of *X. laevis* Y RNA are structurally conserved between the two proteins. His189, Tyr179, Arg148, Arg151, Tyr146 in Rsr are present in similar conformations as in *X. laevis* Ro and possibly

interact with Y RNA in the same manner as seen in *X. laevis* Ro. Some residues that interact with Y RNA in *X. laevis* Ro are not conserved, like residues Lys187 and Asp141. Although the Y RNA substrates of Ro from different species are predicted to form identical secondary structures (111), there are subtle differences in their sequence even within the conserved Ro binding region. It is possible that the non-conserved positions at the Y RNA binding site of Ro account for the diversity in the Ro binding regions of Y RNAs from different species.

Ro in *X. laevis* has also been shown to bind ssRNAs at the central cavity (72),(90). A comparison of this site with Rsr shows that most residues that interact with ssRNA in *X. laevis* Ro, are conserved. Mutagenesis in *X. laevis* Ro has identified the conserved residues corresponding to Lys172, Arg176 to be important in binding ssRNA (90). These along with residues Thr123 and Arg174 interact through hydrogen bonds with the phosphate backbone of the ssRNA. Significantly, residues of *X. laevis* Ro corresponding to the conserved residues Arg273, Arg296 and Arg309 interact with the ssRNA bases but are not critical for binding the ssRNA. In Rsr, these residues lie on the helices H15 to H16 which are displaced away to form a significantly larger central cavity. This suggests that Rsr may bind a variety of RNA substrates in the central cavity by changing the size of the cavity and using the conserved residues in the flexible helices to interact with these RNAs while the primary residues forming the primary, most-stabilizing interactions with the RNA come from the more rigid regions of the protein.

Previous studies in various organisms (68),(69),(72) have shown that besides Y RNAs, Ro also binds U2 snRNAs and pre-5SrRNAs. Also, in *D. radiodurans* at least four different RNAs which associate with Rsr, accumulate post-UV irradiation. Only one of these resembles a Y RNA (83). It is possible that recognition of such a variety of RNAs by Ro is a result of structural adaptation of the protein, made possible by a series of highly conserved motifs which serve as flexible joints in the protein backbone.

The structure of Rsr reveals many conserved Glycine-rich motifs and the positions of these motifs coincide with regions of Rsr which show a marked difference from the previously described structure of Ro from *X. laevis.* These large structural

differences are most pronounced in the region surrounding the central cavity that has been shown to bind ssRNA in *X. laevis*. The outward movement of helices at the interface of the two domains results in a significantly larger central cavity. Other proteins which possess a repeated antiparallel helix-turn-helix or HEAT-repeat domain similar to the N-terminal domain of Rsr, have been shown to be highly flexible and it has been suggested that flexibility may be an intrinsic feature of such domains (112). Also, one of the major changes seen in the Rsr structure is the switch in the conserved β1 region between a beta-strand versus the loop conformation seen in *X. laevis* Ro. Based on the observation that a conserved tyrosine residue from this region forms the second coordination site around the MIDAS metal in the vWFA domain of *X. laevis* Ro but is placed away from the site in Rsr, we speculate that the vWFA domain may play a role in mediating structural changes via this region, all the way to the RNA-binding central cavity.

2.2.7 Rsr: DrY RNA^{35} complex forms large multimers in solution

To further characterize the Rsr:DrY RNA^{35} complex, we performed size exclusion chromatography using a Superose 6 column (range of 5 to 5000 kDa). The elution profiles are shown in **Figure 2.15**. Fractions containing RNA showed a 260/280 ratio greater than 1. Fractions containing protein showed a 260/280 ratio less than 1. Fractions containing the complex of Rsr and RNA showed a 260/280 ration greater than 1 and were detected on a dot-blot stained with Anti-6histidine Antibodies. Details are provided in the methods section. Our results show that Rsr forms multimers upon binding DrY RNA^{35}. To get a more accurate measure of the size of these multimers, similar runs were performed on a Superdex 200 column (**Figure 2.16**). The multimers are of an average size > 700kDa and are composed of ~ 12 molecules of Rsr:Dr Y RNA. Consistent with the tryptophan fluorescence assays described above, the ratio of absorbance 260/280 of the multimers corresponds to a 1:1 complex of Rsr: Y RNA. While the multimer formation occurs in the presence of both Dr Y RNA as well as DrY RNA^{35}, Rsr alone and DrY RNA^{35} alone even over a wide range of concentrations (1μM

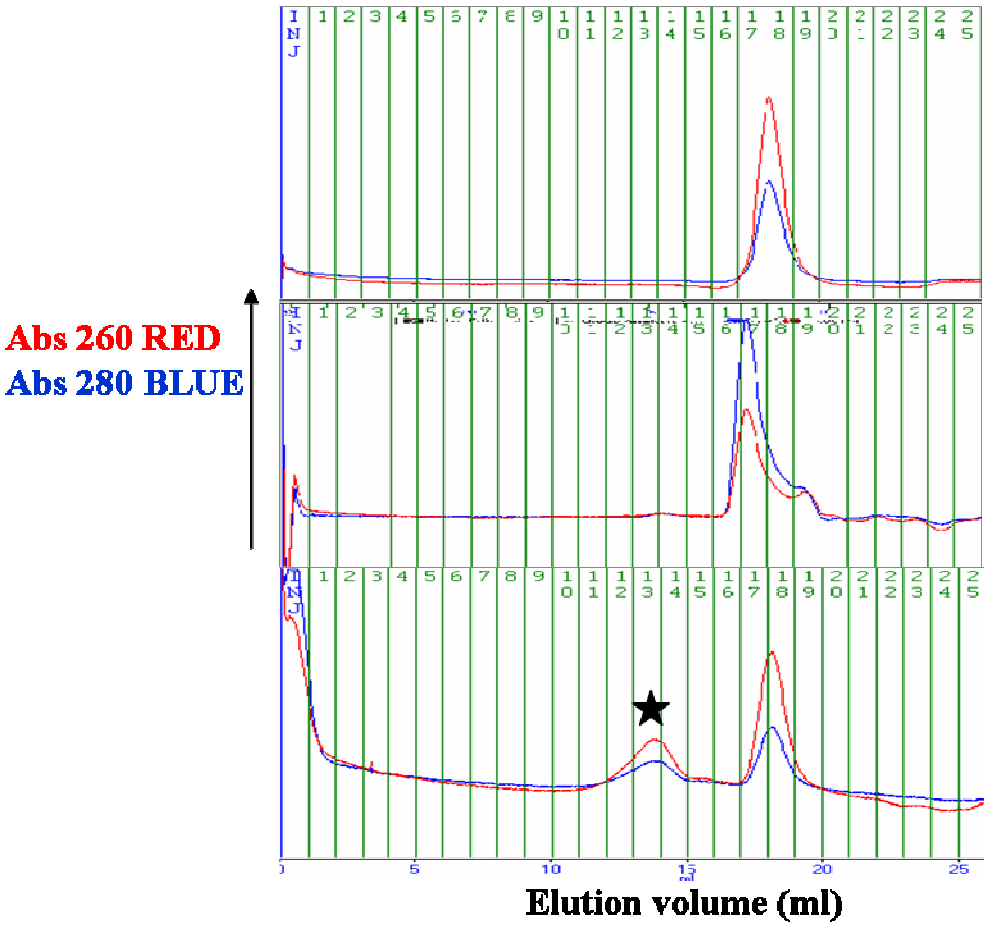


Figure 2.15 Size exclusion profiles from a Superose 6 column

A plot of Absorbance at 280nm (blue) and 260nm (red) versus Elution volume shows the elution characteristics of DrY RNA^{35} alone (top panel), Rsr alone (middle panel) and the Rsr:DrY RNA^{35} complex with excess unbound DrY RNA^{35} (bottom panel). Black asterisk marks the fractions corresponding to multimers of Rsr:DrY RNA^{35} complex.

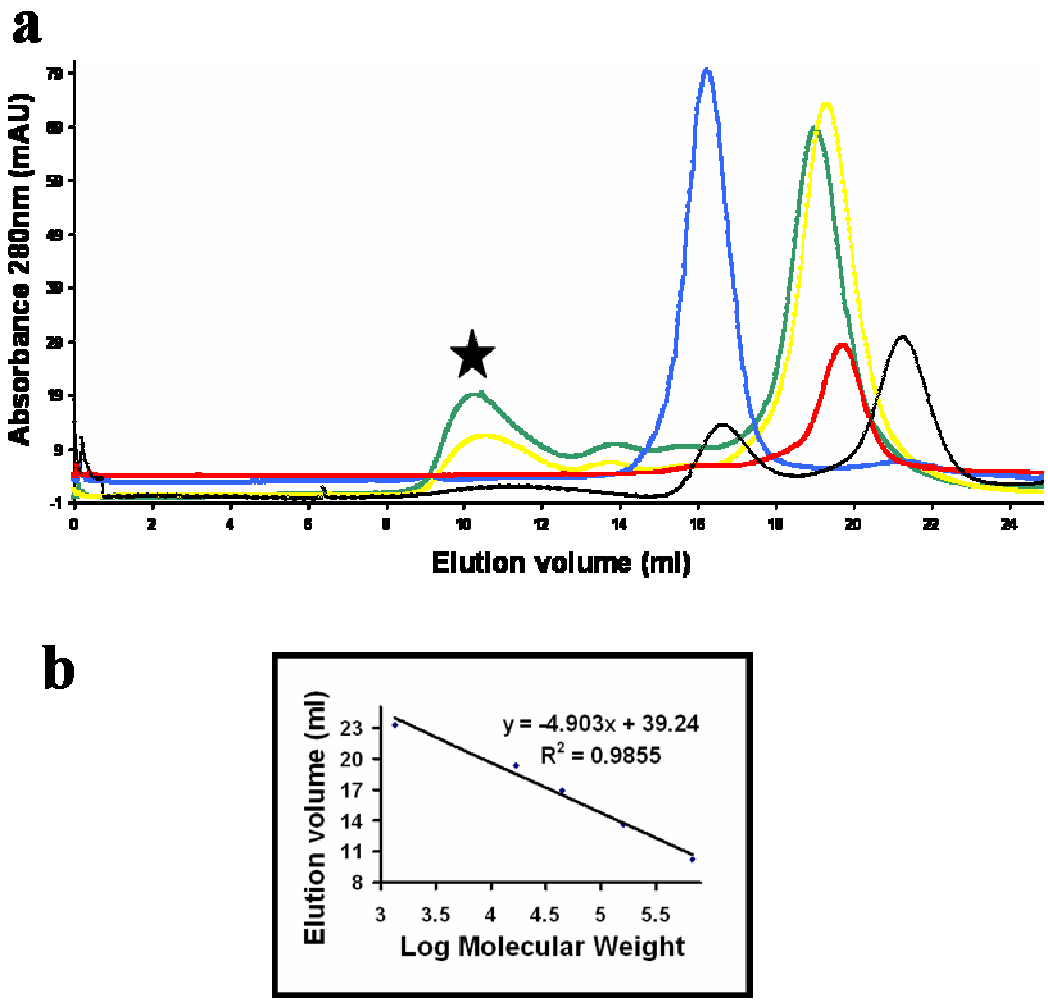


Figure 2.16 Size exclusion profiles from a Superdex-200 column

a. A plot of Absorbance at 280nm versus Elution volume shows the elution characteristics of Rsr alone (blue), DrY RNA^{35} alone (red), the Rsr:DrY RNA^{35} complex and excess unbound DrY RNA^{35} (green) and Rsr:DrY RNA^{35} complex and excess unbound DrY RNA^{35} in buffer supplemented with 10mM metal chelator EDTA (yellow). Black asterisk marks the peak corresponding to multimers of Rsr:DrY RNA^{35} complex. b. Calibration curve obtained from the elution profile of a mix of five standard proteins : Thyroglobulin (670kDa), Bovine gamma globulin (150kDa), Chicken ovalbumin (44kDa), Equine myoglobin (17kDa) and Vitamin B12 (1.3kDa).

to ~500μM) did not show formation of multimers and eluted at the expected monomeric sizes (representative curves are shown in **Figure 2.16**). Similar experiments performed in the presence of saturating amounts of magnesium ion or with an excess of metal chelator EDTA showed identical multimer formation suggesting that the formation of multimers is independent of the presence of metal ions at the vWFA domain of Rsr. Our observations suggest that while the multimers are stable in solution the equilibrium shifts more towards smaller oligomers (for example a 1:1 complex) when the size exclusion column is run at flow rates lower than 0.5ml/min. Dissociation of the multimer also occurs at higher concentrations of salt, which is presumably due to disruption of ionic interactions between Rsr and DrY RNA as seen in the fluorescence assays.

To confirm that the formation of the 700kDa multimer is based on specific interactions and not a consequence of non-specific aggregation, the fractions corresponding to multimers were analyzed by electron microscopy. Electron microscopy was performed by Christos G. Savva and Dr. Andreas Holzenburg at the Microscopy and Imaging Center (MIC) in Texas A&M University. Electron micrographs confirmed the presence of particles displaying symmetrical patterns in projection. Representative class averages obtained from 1600 particles (**Figure 2.17**) suggest 2- to 4- fold rotational symmetry residing within the multimers measuring up to 180Å along the long axis. In addition, some classes did not reveal any rotational symmetry in projection. The observed heterogeneity within the Rsr:DrY RNA population has so far hampered 3-dimensional reconstruction efforts.

Since the formation of multimers occurs in the presence of both the full length Y RNA as well as the conserved region of Y RNA and the fluorescence assays suggest a 1:1 complex between Rsr and Y RNA, it is likely that the multimer formation does not involve additional RNA-RNA or RNA-Protein interactions. This suggests that upon Y RNA binding, Rsr adopts an alternate conformation which promotes oligomerization. This feature of forming large, ordered multimers in the presence of Y RNA has not been reported yet for other Ro orthologs. Formation of higher-order oligomers upon binding single-stranded DNA has been observed for the Human RecQ helicase (113). The

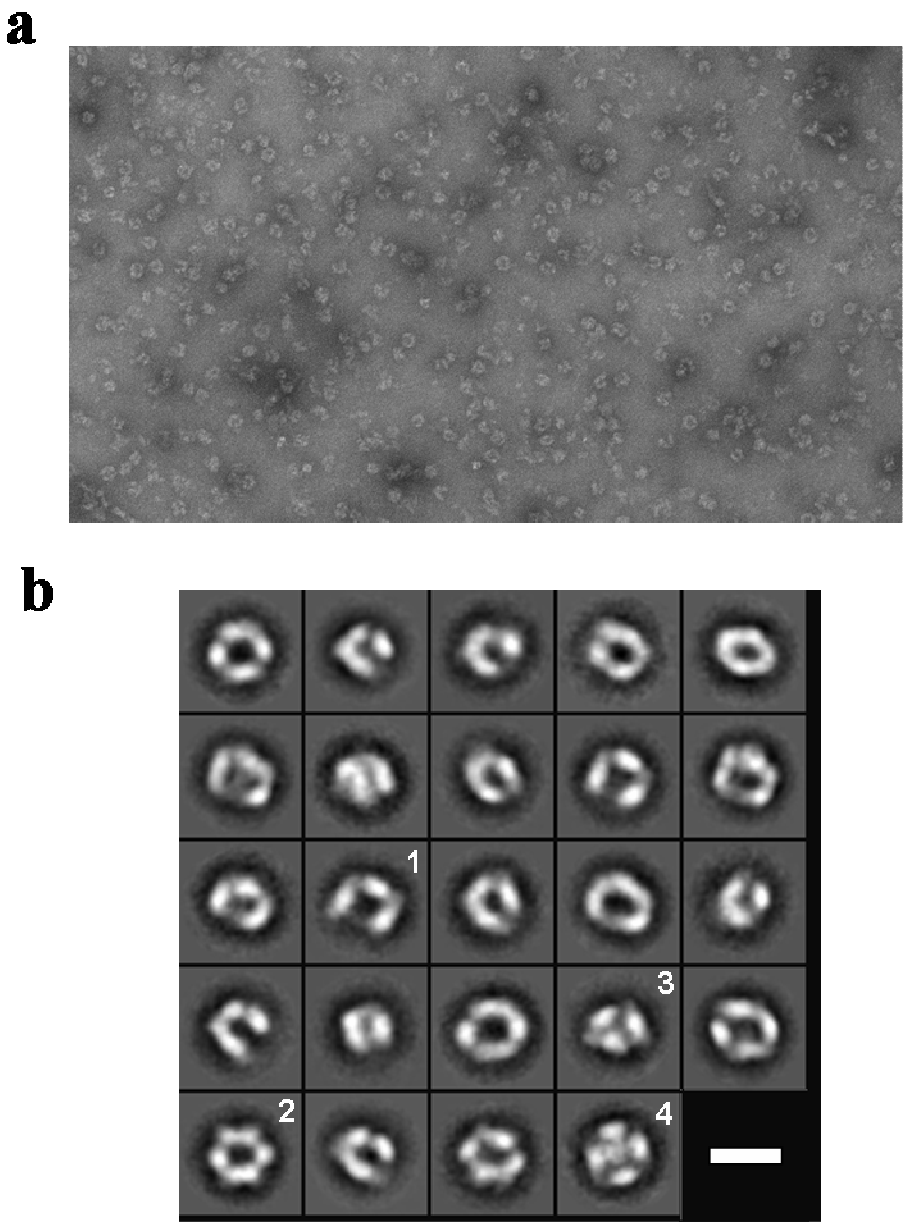


Figure 2.17 Electron microscopic analysis of single Rsr:DrY RNA[35] particles

a. Electron micrograph of negatively stained Rsr:DrY RNA multimers shows ringlike formations. b. Representative class averages displaying no symmetry (1), 2-fold (2), 3 and 4-fold (3, 4) rotational symmetry. Scale bar corresponds to 20 nm.

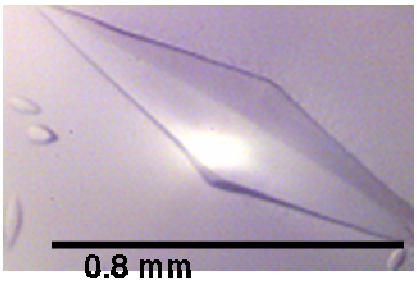
RecQ-ssDNA complex forms oligomeric ringlike structures surrounding a central pore, the site where annealing of complementary strands of DNA could take place. This higher-order oligomerization in the presence of ssDNA has been shown to regulate the enzyme, switching the activity from DNA unwinding to DNA strand annealing. Although the RNA binding properties of Ro have been extensively characterized (72), it is not known what events take place after Ro binds Y RNA and what triggers those events to take place. It is possible that formation of these multimers acts as a trigger to set off the downstream processing events, which in the case of *D. radiodurans* is likely to be the removal of UV damaged RNA. Human RoRNPs from pull-down assays using Anti-Ro antibodies are shown to form ~320kDa sized particles (106).

2.2.8 Crystallization and structure determination of Rsr:DrY RNA complex

Although complete only to low resolution, the diffraction data obtained from crystals of the Rsr:DrY RNA complex (**Figure 2.18**) was used for Molecular replacement. Initially the Rsr structure and eventually the RNA bound structure of *X. laevis* Ro were used as search models for the molecular replacement runs. One of the major complications in these runs was the correct estimation of the number of molecules in the asymmetric unit of the Rsr:DrY RNA complex crystals. Calculations using the Mathews Coefficient program in the ccp4 suite suggested anywhere between 10 to 14 molecules in the asymmetric unit (**Figure 2.18**), adding another level of complexity to the MR runs. Phaser, MolRep and Amore failed to find solutions when 12 molecules were searched for. To improve the chances of a successful MR run, the quality of diffraction data needs to be improved.

Crystals of the complex also grew in the presence of Seleno-methionine substituted Rsr and diffracted up to 7Å. Three wavelength MAD data was collected and used for phase calculation using phenix.hyss and Solve. No heavy atom sites were located by either program. Inherent disorder in the crystals due to heterogeneity of the sample, handling of the crystals, packing in the crystal lattice could be some of the reasons for the poor quality diffraction.

a



b

Nmol/ASU	Mathews Coefficient	% solvent
1	32.69	96.24
10	3.27	62.4
11	2.97	58.6
12	2.72	54.8
13	2.51	51.1
14	2.33	47.3

Figure 2.18 Crystallographic analysis of Rsr:DrY RNA[35] complex

a. Crystals of 6hisRsr:DrY RNA[35] grown in 30% Methyl Pentanediol, 0.2M Ammonium acetate, 0.1M Sodium citrate pH5.6. b. Calculation of Mathew's coefficient searching for a 77000 Dalton molecule in the Rsr:DrY RNA[35] unit cell suggests ~ 12 to 14 molecules in the asymmetric unit.

2.3 EXPERIMENTAL APPROACH AND DESIGN

2.3.1 Construction of an *E. coli* overexpression plasmid for Rsr and purification of Rsr

The Rsr coding region was amplified by PCR from *D. radiodurans* genomic DNA (purchased from ATCC) and cloned into pET28b (Novagen). The primers and restriction sites used are listed in **Table 2.4**. The resultant plasmid was transformed into *E. coli* BL21 (DE3) and grown in LB media and 50μg/ml of kanamycin to mid-log phase. IPTG was then added to a final concentration of 0.5mM and the cells were allowed to grow for ~18h at 25ºC. The temperature of 25ºC for cell growth after IPTG induction was obtained after analytical experiments were performed at 37 ºC, 25 ºC and 16 ºC. At 37ºC Rsr was expressed as insoluble aggregates where as at 16 ºC there was no expression. Small scale analytical experiments were also done to find the *E. coli* strain best suited for Rsr over-expression. Rsr was over expressed in both BL21 (DE3) and Rosetta (DE3) cells but not in plysS^{+} strains.

2.3.2 Cloning of the TROVE domain of Rsr

The N-terminal residues (#24 to 360) were amplified by PCR and cloned into different vectors to obtain N-terminal and C-terminal 6-histidine tagged constructs. The primers, vectors and restriction sites used are listed in **Table 2.4**. Based on the structure of Rsr, since the N-terminal 35 residues were disordered, a second construct of the TROVE domain was made, comprising residues 36 to 360 to help in making the protein soluble. This however did not help and the construct was still expressed as inclusion bodies in *E. coli*.

Table 2.4 Cloning strategy for various Rsr and DrY RNA constructs

Construct	Vector	Restriction sites	Primers
6h-Rsr	pET28b	Nde1/Xho1	5'ctagctagcatgaagaacttgctccgtgcc 5'ccgctcgagtcaaacc tcgcccccgcgcaaaagccg
6h-K172A	pET28b	Nde1/Xho1	5'caagctggcgctgtgggccgtggcgtacaaggcccgcgacggctgg
6h-K176A	pET28b	Nde1/Xho1	5'gtgggccgtggcgtacaaggccgccgacggctggtcccaggccgac
6h-H189S	pET28b	Nde1/Xho1	5'gccgacgcgctgcgtaaggccagcccaagaccgacgacgcggcc
6h-G267A	pET28b	Nde1/Xho1	5'gcgccgccatgcagaccaacgccctaacctggctgctgcgcaa
6h-TROVE	pET28b	Nde1/Xho1	5'gggaattccatatggtcaggaacaacgcgggcggcttcg 5'ccgctcgagttaggtgttcgccggctgcacgttgcc
TROVE-6h	pET28b	Nde/Xho1	5'gggaattccatatggtcaggaacaacgcgggcggcttcg 5'ccgctcgagggtgttcgccggctgcacgttgcc
6h-TROVE	pSV281	BamH1/Xho1	5'cgcggatccgtcaggaacaacgcgggcggcttcg 5'ccgctcgagttaggtgttcgccggctgcacgttgcc
6h-TROVE^{-35}	pET28b	Nde1/Xho1	5'ctagctagcgacgagtctcgcctgacccgg 5'ccgctcgagttaggtgttcgccggctgcacgttgcc
6h-Rsr^{-35}	pET28b	Nde1/Xho1	5'ctagctagcgacgagtctcgcctgacccgg 5'ccgctcgagtcaaacctcgcccccgcgcaaaagccg
DrY RNA^{123}	pETBluel	Blunt	5' atttaggtgacactatagaagctctgggccg 5' agcgctctggacaagggttcgggtcacgg
DrY RNA^{123}	pETBluc1	Blunt	5' atttaggtgacactatagacgctctgggccg 5' gttgacgctctggacaagggttcgggtcacgg

2.3.3 Purification of Rsr

Cells were lysed using the French press, in 25mMHepes, 150mMNaCl, 2-βME, 1X Protease Inhibitor cocktail SetV (Calbiochem). After centrifugation at 15000Xg for 60 minutes, the supernatant subjected to Ni^{2+} affinity chromatography. Following washes with buffer containing Hepes, NaCl, 2 βME and the buffer supplemented with 20mM Imidazole, elution was carried out using a linear gradient of Imidazole ranging from 20mM and 400mM in the buffer. Fractions corresponding to approximately 100mM Imidazole contained Rsr, the purity of which was greater than 90% as judged by Coomassie-stained 4-20% Tris-HCl SDS-PAGE gels. The protein was concentrated using Amicon-Ultra centrifugal filters- MWCO 10kDa. Initial Crystal screens performed at a protein concentration of 1mg/ml and 3mg/ml showed that only drops containing phosphates/sulfates/cacodylate in the crystallization condition were clear, suggesting that phosphates/sulfates/cacodylate may be necessary to keep the protein soluble. To identify the most suitable buffer to maintain Rsr in a stable and soluble form, small-scale dialysis was performed against different buffers (buffer1: 50mM Phosphate 7.0, 150mM NaCl, 2mM DTT, buffer2: 500mM Phosphate 7.0, 150mM NaCl, 2DTT, buffer 3: 50mM Phosphate 7.0, 500mM NaCl, 2DTT, buffer4: 100mM Cacodylate 7.0, 150mM NaCl, 2mM DTT, buffer5: 500mM Cacodylate 7.0, 150mM NaCl, 2DTT) with 3ml of Rsr(at 2mg/ml concentration) in dialysis cassettes. Since Rsr stayed most soluble in buffer containing 50mM Phosphate 7.0, 500mM NaCl and 2mM DTT, these components were used in subsequent purifications of Rsr.

Cells were lysed in buffer A (25 mM Phosphate pH 7.0, 500mM Sodium chloride and 2 mM βME) by french press. The lysate was centrifuged at 15000rpm and the supernatant subjected to Ni^{2+} affinity chromatography. Following washes with buffer A (~ 20X column volume) and buffer A supplemented with 20mM Imidazole (30X column volume), elution was carried out using a linear gradient of imidazole ranging from 20mM and 400mM in buffer A over 30X column volume. Fractions corresponding to ~100mM Imidazole contained Rsr, the purity of which was greater than 90% as judged by Coomassie-stained 4-20% Tris-HCl SDS-PAGE gels (**Figure 2.19**). Fractions

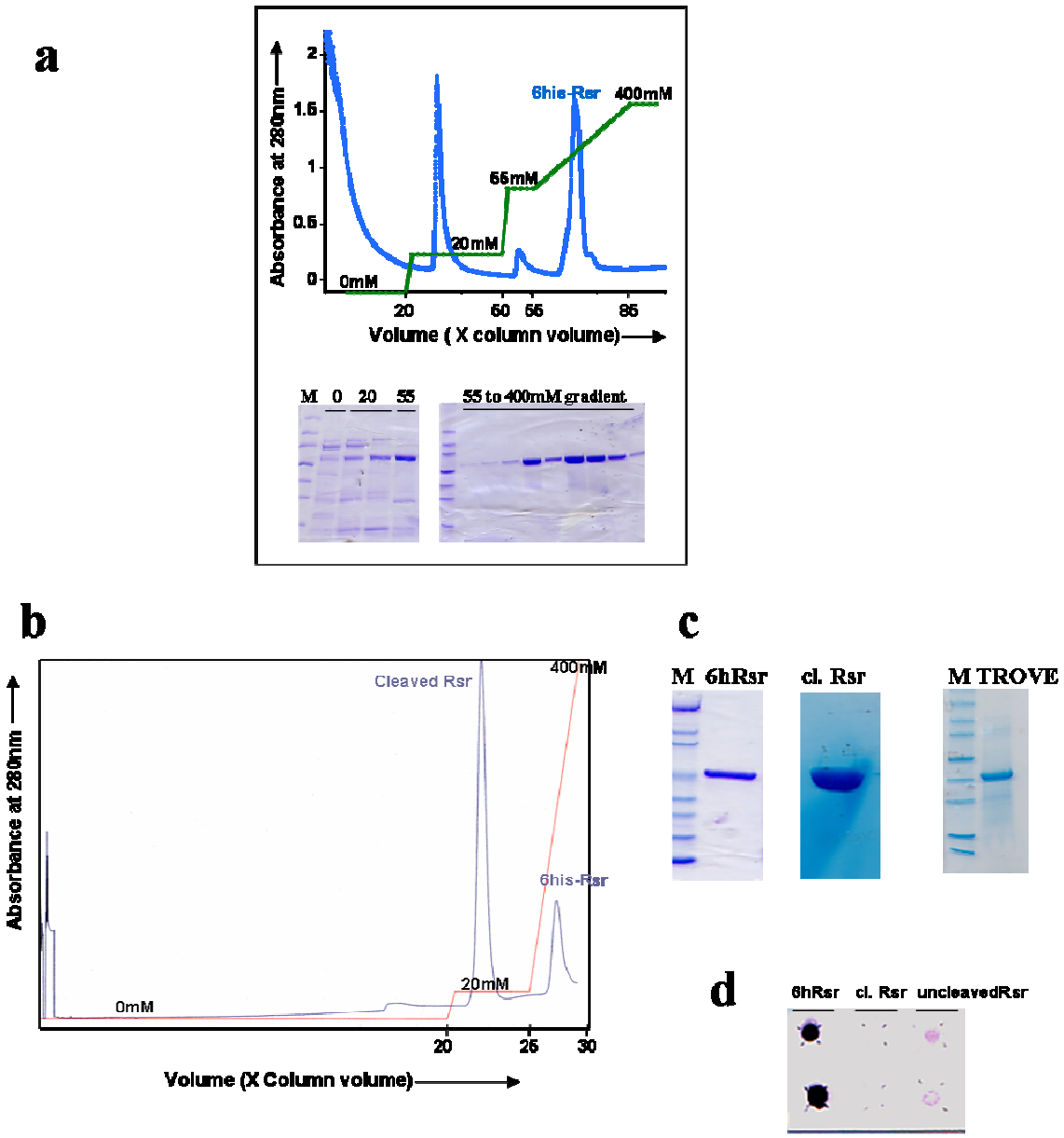


Figure 2.19 Purification of Rsr

a. Plot of absorbance at 280nm versus volume shows the elution profile of Rsr from a Nickel column. Steps corresponding to increasing concentrations of Imidazole (green) are shown. Pure Rsr elutes at >100mM Imidazole. SDS-PAGE analysis of the fractions confirms the presence of Rsr in the fractions. b. Elution profile post-Thrombin cleavage of 6his-Rsr is shown. Cleaved Rsr elutes at 20mM Imidazole whereas uncleaved Rsr elutes at ~100mM Imidazole. c. SDS-PAGE analysis of samples used for crystallization shows >90% pure 6his-Rsr, cleaved Rsr and ~80% pure 6his-TROVE. D. Dot blot stained with Anti-6his Antibodies confirms the purity of the samples of cleaved and uncleaved Rsr.

containing Rsr were cleaved with Thrombin (purchased from Calbiochem) at a concentration of 5 Units of Thrombin per milligram of Rsr during dialysis against 25mM Phosphate pH 7.0, 150mM sodium chloride and 2mM βME. The dialysed sample contained Rsr with and without the 6-histidine tag and these were separated again by Ni^{2+} affinity chromatography, with Rsr being eluted with 20mM imidazole (**Figure 2.19**). This fraction was concentrated using amicon centrifugal filters and buffer exchange with 25mM Phosphate pH 7.0, 150mM NaCl and 2mM DTT was done during concentration. All mutant proteins were purified the same way. Protein concentration was calculated using absorbance at 280nm and extinction coefficient equal to 57660 $M^{-1}cm^{-1}$ (114). To analyze its oligomeric state, Rsr was subjected to size exclusion chromatography on a Superdex-S200 column. Rsr migrated as a single monomeric species on the column when compared to the standard calibration curve (**Figure 2.16**).

To avoid precipitation of the protein which accompanies lengthy purification protocols with multiple chromatography steps, the first step of Nickel affinity purification of 6his Rsr was optimized to obtain >90% purity of the sample. This was more important in this case since the purification post-Thrombin cleavage resulted in the desired protein being eluted in the lower concentrations of eluent (Imidazole). Optimization of the 1^{st} purification step was done by introducing longer wash steps at 0mM, 20mM Imidazole and an extra 55mM Imidazole step.

2.3.4 Use of Crystallization screens to find the optimal buffer for Rsr

Initial purification of Rsr in 25mM Hepes pH7.0, 250mMNaCl and 2mM βME on a Ni^{2+} affinity column resulted in pure protein which would slowly precipitate over a period of 12 hours. This pure protein could not be concentrated beyond 3mg/ml. This hindered crystal growth since one of the requirements of Crystal formation is that the sample be stable enough to last the period of crystal growth. Preliminary screens setup with Rsr at 2mg/ml in 25mM Hepes pH7.0, 250mM NaCl, 2mM βME resulted in 95% of the drops forming brown precipitation within a day. The remaining 5% of the drops remained clear for more than 5 days. The common factor among these 5% of clear conditions was

the presence of Phosphate/ Sulfate/ Cacodylate salts. Sulfate and Cacodylate share similar chemical properties as Phosphate thereby behaving like Phosphate mimics, often present in crystals in place of the physiological relevant Phosphate ion. Dialysis of purified Rsr for 3 days against different solutions containing Phosphate/ Sulfate and Cacodylate with varying amounts of Sodium Chloride resulted in clear protein solution only in the presence of 25mM Phosphate pH7.0, 150mM NaCl and 2mM βME. In this buffer, Rsr could be concentrated to >25mg/ml and remained stable for more than 1 week at crystallization temperatures (18ºC), hence this buffer was used to purify the protein for subsequent crystallization experiments. The requirement of Phosphate salts to stabilize purified protein has been shown for other RNA binding proteins. This approach of using crystallization screens to find the optimal buffer for the stability of a protein has broad general applications.

2.3.5 Expression and purification of TROVE domain

All constructs of the TROVE domain were expressed in the Rosetta(DE3) plysS strain of *E. coli*. The cells were lysed using French press in buffer containing 25mM phosphate pH 7.0, 500mM NaCl and 2mM βME. Following lysis and a clearing spin at 3000Xg, the supernatant was centrifuged at 10000Xg for 20 minutes. Partially purified TROVE protein was present in the pellet as inclusion bodies. Since all constructs were expressed as insoluble aggregates, the N-terminal 6his tagged protein was used for subsequent experiments. The inclusion bodies were solubilized in 0.5% Empigen BB, a zwitterionic detergent. Following solubilization, the sample was loaded on a Ni^{2+} affinity column and eluted with buffer containing ~100mM Imidazole. The fractions containing pure 6-his TROVE were dialysed against 25mM phosphate pH 7.0, 150mM NaCl, 2mM DTT. To facilitate removal of Empigen BB, Bio-beads (Calbiochem) were added to the dialysis buffer and many buffer exchanges were done, each supplemented with fresh bio-beads.

Since the predicted TROVE domain is expressed as inclusion bodies in *E. coli* cells, Trypsin digestion of Rsr was performed, to identify a domain in Rsr that could be

expressed in the soluble fraction. SDS-PAGE analysis of Trypsin digested Rsr shows at least two prominent domains (**Figure 2.20**). The major band on the gel migrates smaller than the predicted TROVE domain, and it is likely that cloning this region and expressing it in *E. coli* could be a good starting point to study the domains of Rsr individually.

2.3.6 Crystallization of Rsr

Crystals are ordered arrays of the sample protein/macromolecule arranged in a lattice (115). The crystalline state is described as a metastable state between the super-saturated solution and precipitate or insoluble states (115), (116). Hence, crystallization is more likely when the starting sample used is at a concentration close to super saturation in the given buffer. The process of crystallization involves the precipitation of the sample at a very controlled and slow rate so that along the path of precipitation, the sample could form crystals and be trapped in the metastable crystalline state (115). Consistent with this, commonly used crystallization agents are typically precipitation agents like high concentrations of salt (ammonium sulfate, sodium phosphate, sodium malonate), hygroscopic polymers (Polyethylene glycol, polyethylene imine) and alcohols (Methyl pentane-diol, ethanol) (115). In many cases, crystallization is also caused by varying the pH of the solution, driving it towards the pI of the protein where the protein tends to be most unstable (115). The important step in the crystallization experiment is to control the process so as to not remain in the completely stable and soluble state indefinitely, but not be driven towards precipitation too rapidly.

Initial screening for Rsr crystals was done using the 6-histine tagged protein at a concentration of 10mg/ml. Two types of crystals were obtained. Microcrystals (**Figure 2.2**) grew in conditions containing Polyethylene glycol as the precipitant and cube shaped crystals were obtained by the sitting drop method in Intelli plates, in 1.14M Sodium-Potassium tartrate and 0.1MHepes pH 7.5 (Magnesium free screen 1, condition #52, Lawrence Livermore, Berkeley). These crystals could not be reproduced in home-made conditions or fresh solution supplied by the Lawrence Livermore Laboratory. To

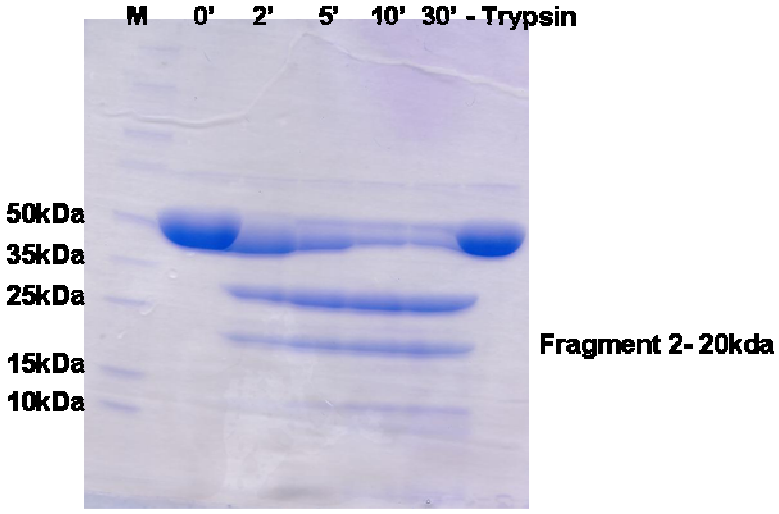


Figure 2.20 Trypsin digestion of Rsr

Time course of Trypsin digestion of Rsr shows two prominent domains, one of approximately 30kDa and the other around 20kDa in molecular weight.

confirm that the crystallized protein was intact 6-histidine Rsr, the crystal was dissolved in SDS-PAGE loading buffer. The sample was run on an SDS-PAGE, followed by Western blotting on a nitrocellulose membrane and detected with Anti-6his-Antibodies. Since the crystals with 6-histidine Rsr were irreproducible, screening for crystals was done post cleavage with Thrombin protease. Rod shaped crystals for native Rsr without the 6-his tag were grown by the sitting drop method in Intelli plates, in a crystallization solution consisting of 0.1 M imidazole pH 8.0, 12% polyethylene glycol 8000 and 0.2M calcium acetate. Diffraction quality crystals (**Figure 2.2**) were obtained by mixing 4 µL of crystallization solution with 2 µL protein at a concentration of 400 µM, and equilibrating against 100 µL of crystallization solution.

To obtain phase information for structure determination, seleno-methionine (Se-met) substituted Rsr was produced in *E. coli* B834 (DE3) cells. Briefly, the cells were transformed with the pET28b-Rsr construct, grown in LB media to mid-log phase, transferred to M9 minimal media supplemented with L-amino acids (Sigma, 50 mg/L), L-seleno-methionine (Acros, 50 mg/L), glucose (0.4% w/v), calcium chloride 0.1 mM), magnesium sulfate (4 mM), thiamine (5 µg/mL) and kanamycin (50 µg/mL), induced with 0.5 mM IPTG and allowed to grow for 18 h at 25 ˚C. Following harvesting of culture, the same purification protocol was followed as for the native Rsr except increasing the concentration of DTT from 2mM to 5mM in the final buffer prior to crystallization. Crystals for Se-met substituted Rsr were obtained in the same condition as that of native Rsr.

2.3.7 Preparing crystals for data collection

To minimize the harmful effects of the penetrating X-ray beam on the crystal, data collection is typically performed with the crystal placed in a stream of liquid nitrogen, at a temperature of 100K (117). To prevent freezing of the solvent in the crystals and subsequent fracturing of the crystals, the crystals are picked using inert nylon loops and soaked for 1-30 seconds in a solution of cryoprotectant which prevents freezing (115). This is an important optimization step, often involving trial and error since many crystals

do not survive or do not diffract in commonly used cryoprotectants like 10 to 40% glycerol, 10 to 40% Ethylene glycol, Polyethylene Glycol etc. Some cryoprotectants are oil based compounds like paraffin oil, silicone oil, FOMBLIN, and the highly viscous paratone (115). The viscosity of the cryoprotectants often causes mechanical damage to fragile crystals. The few crystals of 6his Rsr obtained in the initial screen were flash frozen in liquid nitrogen using a mix of 70% paratone and 30% PCR Mineral oil as cryoprotectant. This home-made mix of two oil based compounds served as a good cryoprotectant as well as decreased the viscosity of the paratone which was necessary to maintain the crystals intact. The crystals of cleaved Rsr were flash frozen in liquid nitrogen using 25% glycerol in mother liquor as a cryoprotectant. The Se-met Rsr crystals were flash frozen in liquid nitrogen using paratone as the cryo-protectant. The solution of 25% glycerol in mother liquor that was used successfully for the native Rsr crystals did not yield good diffraction for the Se-met crystals.

2.3.8 X-ray data collection and data processing

X-rays are penetrating electromagnetic radiation in the 0.5Å to 100nm range (117). Since the atomic bond distances in biological macromolecules are of the order of 1 to 3Å, the wavelength of X-rays is ideal to resolve atoms separated by similar distances (117). X-rays interacting with matter get absorbed to result in vibration of electrons with the same frequency as the incident waves (117). The vibrating electrons, when returning to the ground state emit X- radiation similar in energy and wavelength to the incident radiation.

X-ray diffraction is a phenomenon when X-rays are scattered from a particle in different directions, such that the scattered rays have the same wavelength as the incident rays (117), (118). Scattered X-rays from a regular arrangement of scattering particles undergo constructive interference along some directions. The X-rays scattered in particular directions are called orders of diffraction (h), with the direction closest to the unscattered ray being the first order of diffraction (117). If there are enough particles repeating in a lattice (as in a crystal), the emitted X-rays undergo substantial constructive

interference, resulting in a proportional increase in the intensity/amplitude of the X-ray. When sufficient intensity has been gained, the X-ray makes a characteristic diffraction spot on a detector like photographic film, or modern day detectors like charge-couple devices etc. (117). For every such diffracted X-ray that produces a diffraction spot, there is an associated amplitude and phase angle (117), (118). The amplitude and phase angle for the diffracted X-rays represent the Fourier transform of the scattering density (electrons, in this case) (117). Fourier transform is a mathematical expression used to describe the harmonics of any repeating function, for example the repeating density of electrons from identical molecules that make up a crystal (117). All the scattering that contributes to the diffraction in a given direction can be can be summed up, to obtain a quantity called the Structure Factor, F_h which is a complex number with an amplitude A (= $|F_h|$) and phase (= α_h) (117).

$$F_h = |F_h| \exp[i\alpha_h] = {}_0\int\rho(x) \exp [2\pi i h x] \, dx$$

Similarly, $\rho(x)$ can be derived from the inverse Fourier transform summed over all the scattered X-rays, if the scattering factor F_h is known (117). $\rho(x)$ represents the repeating factor in the crystal, which is the density of scatterers or the electron density. In the X-ray diffraction experiment, the measurable quantity is the intensity of the scattered X-ray. Intensity is proportional to the square of the amplitude, so the amplitude of the diffracted X-ray can be calculated (117). However, information about the phase angle of the X-rays is lost in the diffraction experiment, giving rise to the *phase problem* in crystallography. Phases, therefore, have to be calculated using other methods, some of which are described in the following section. In the diffraction experiment, the greater the change in direction of the scattered rays from the direction of the incident rays, the finer the spacing that can be resolved between the scatterers (117), (118). The spacing between the scatterers is called *d-spacing* and is related to a parameter called *resolution*. Resolution is expressed in Angstroms since spacing in between scatteres in macromolecular crystals is of that order (118). At a given wavelength, with n being an integral value, resolution (d) is related to the scattering angle (θ) by Bragg's Law (118):

$$2d \sin\theta = n\lambda$$

The repeating arrangement of molecules in the crystal lattice gives rise to symmetry. Two objects are related by symmetry if one molecule is subjected to an operation (like rotation, inversion, mirror image formation) and as a result is superimposable on the other molecule. Restrictions due to the chirality of amino-acids allow only rotational operations to exist in protein crystals (117). In 3-dimensional crystals, different rotational symmetries could exist along the x, y and z axis. Constraints of the crystal lattice allow only 2-, 3-, 4- and 6- fold rotational symmetries to exist in macromolecular crystals and the possible kinds of symmetry that can be possessed by crystals are described together as the *crystallographic point-groups* (119). Besides the rotational symmetry, crystals also possess a symmetry called the *lattice translational symmetry* generated by repetition of the macromolecule along 3 different axes (117). Lattice translation along with rotational symmetry gives rise to complex symmetries within crystals. A combination of rotational and translational symmetries gives rise to *screw* symmetry (117). The three lattice translations, along the x, y and z axis define the *Unit Cell* (117). The Unit Cell can be imagined to be a block which when translated along the x, y and z directions gives rise to the entire crystal lattice. The 3 lattice translations are defined as *a*, *b* and *c* and the angles between them are defined as α, β and γ. The different types of Unit Cells found in crystals, have been grouped together to describe *7 Crystal Systems* (117). Calculation of the symmetry in the crystal is a necessary and important step in structure determination. The diffraction pattern obtained from a crystal reflects on the symmetry present in the crystal, hence looking at the pattern of diffraction spots, one can calculate the inherent symmetry within the crystal (118). This can be done manually, by looking at the precession frames of the diffraction data or using programs that are trained with the known rules of symmetry calculation (118). The minimum unit beyond which no symmetry exists is called the *Asymmetric Unit* (118). It is enough to solve the structure of the asymmetric unit since the rest of the crystal can be generated by applying the correct symmetry.

The 6-his Rsr crystals were exposed to synchrotron X-rays at beamline19ID, Argonne National Labs. The best crystal (**Figure 2.2**) diffracted to ~5Å. The best

crystal of native Rsr diffracted to 1.89Å at beamline 23-ID while the best crystal of Se-met Rsr diffracted to 2.6 Å at beamline 23-ID at the Argonne Labs, IL.

Rsr crystals showed C2 symmetry belonging to the *monoclinic* crystal system (119). The monoclinic system is characterized by a minimum of one 2-fold symmetry axis. Conventionally, b is parallel to the 2-fold axis and a and c are perpendicular to the 2-fold axis (119). α and $\gamma = 90^{\circ}$ and $\beta \neq 90^{\circ}$. The C2 space group is a lattice that is face centered on the face whose edges are a and b (119). The equivalent points or symmetry related points in the unit cell are: (x, y, z), (-x,y,-z), (1/2+x,1/2+y,z), (1/2-x,1/2+y,-z). Rsr crystals belong to the conventional monoclinic space group with unit cell dimensions of a = 103.56 Å, b = 87.61 Å, and c = 70.62 Å, $\alpha = 90.0^{\circ}$, $\beta = 96.57^{\circ}$, $\gamma = 90.0^{\circ}$. The Se-met Rsr crystals possess C2 symmetry and the data was isomorphous to the native Rsr diffraction data, with cell dimensions of a =105.41Å, b =89.39Å, and c =70.55Å, $\alpha = 90.0^{\circ}$, $\beta = 96.8^{\circ}$, $\gamma = 90.0^{\circ}$. Data statistics are compiled in **Table 2.1**. Completeness of the Rsr data, defined as ratio of the number of unique relections measured to the total number of unique reflections, was equal to 98% for the overall data and 85.3% at a resolution of 1.89Å. A single molecule of Rsr was calculated to compose the Asymmetric Unit. The Mathews volume, which represents the volume of the crystal per unit mass of the protein, was calculated as 2.7 with a solvent content of 54.8%. This is typical for most proteins, the range being 1.9 to 4.2.

2.3.9 Structure determination

As described above, the scattered X-rays from a crystal possess the same energy as the incident X-rays and are in phase with the incident rays (117). However, when certain heavy atoms (with unusually large number of electrons) are present in the crystal, if the wavelength of incident light is close to the transition energy of the heavy atom, then the vibration of the electrons goes out of step and the scattered X-rays are out-of-phase with the incident rays (117). This phenomenon is called *anomalous scattering*. Due to this out-of-phase scattering, the normal condition of $F_{h,k,l}$ and $F_{-h,-k,-l}$ being equal in magnitude and opposite in phase (called *Friedel's Law*) is violated (117). In anomalous

scattering, the Friedel pairs $F_{h,k,l}$ and $F_{-h,-k,-l}$ are not opposite in phase. The difference between $|F_{hkl}|$ and $|F_{-h,-k,-l}|$ is called the *Bijvoet amplitude difference* (117), (118). At the wavelength close to the absorption edge of transition for the heavy atom, the heavy atom displays anomalous scattering but at wavelengths far from the absorption edge, the heavy atom displays normal scattering (118).

Tunable X-ray sources like synchrotrons allow crystals to be exposed to different wavelengths close to and far from the absorption edge of the heavy atom, therefore allowing data collection to determine phases using *Multiwavelength Anomalous Dispersion* (MAD) (118). This MAD technique was used to solve the phase problem for Rsr. The 3 wavelengths used are listed in **Table 2.1**. The peak and inflection wavelengths are at and close to the absorption edge for selenium and the remote wavelength is far away from the edge. From the MAD data, the Bijvoet amplitude differences are calculated, the squares of which are used as coefficients to calculate the *Patterson function* (118). The Patterson function is a Fourier transform which is described using only intensities and is independent of phases. For a small number of scatterers, the Patterson function can be directly calculated from the observed intensities. Since the number of heavy atoms in the crystal is much smaller than the total number of atoms, the Patterson function is used to locate the positions of the heavy atoms (118). On the Patterson map, the positions between pairs of anomalous scatterers show up as cross-peaks.

Another method for phase calculation is called *Molecular Replacement* (117) which relies on an already available homologous structure, called the *search model.* Based on the assumption that there is enough structural similarity between the search model and the unknown structure, the first step is to perform rotational and translational operations on the search model to fit it into the Unit Cell of the unknown protein. For this, an accurate calculation of the Unit Cell of the unknown protein is essential, along with an accurate measure of the contents of the asymmetric unit. For example, if 2 molecules are calculated to be present in the asymmetric unit of the unknown protein, then 2 molecules of the search model need to be placed in the Unit Cell of the unknown

protein. Once an operation has been found that places the search model in the Unit Cell of the unknown protein, the operation is applied to every atom of the search model to obtain new coordinates which represent an initial calculated model for the unknown protein. The new coordinates can be described as: $x_X=Rx_M+T$, where R is the rotational operation and T is the translational operation, x_X denotes the coordinates for the unknown protein and x_M denotes the coordinates of the search model (117). The new coordinates are used to calculate structure factors with the observed intensities (since those are obtained from the diffraction experiment) and the calculated phases. The model of the unknown protein at this stage is biased towards the search model. Bias minimization is done till features exclusive to the unknown protein and not present in the search model can be identified in the electron density. At every step, comparison is made between the amplitudes of the calculated and observed structure factors. As the new model approaches the correct structure, the calculated phases improve to give a less biased and more reliable structure.

After the electron density has been calculated, model building is done using the known rules of protein structures like bond lengths, types of amino acids, planarity of the peptide bond etc. Automated model building programs are trained on these known parameters and can be given electron density as input to build the model. One such model building program called TEXTAL (120) was used for Rsr. TEXTAL uses a 4-step process for model building. First, it locates the model in the correct space in the electron density and then builds the Cα chain using CAPRA (C-Alpha Pattern Recognition Algorithm). In the third step, the side chains are placed in density according to the sequence of the protein which is provided as input for the program. In the fourth step, refinement is done to match the model with the calculated electron density. The accuracy of model building by TEXTAL greatly increases if the data (and hence electron density) provided as input are in the same resolution range as the structures that constitute its training set. After the initial step of automated model building, corrections to the model can be made either manually or using programs like the *Iterative Build* feature in the Phenix package (121). For Rsr, both methods were

employed to obtain the best results.

Structural refinement involves adjustment of the structure to best match the observed crystallographic data, i.e. the amplitudes of the structure factors. Typically the model is refined for the coordinates, the *B values* for every atom and the occupancies. The B- value is a measurement of the displacement of an atom due to thermal motion or disorder (117), (118). For successful refinement, enough good quality data is required along with a model that agrees reasonably well with the electron density. Refinement is usually done in stages, starting with *rigid body refinement,* where the coordinates are moved as one or more rigid bodies to fit better into the electron density (118). Known structural information like constraints in the backbone angles, planar side chains, steric clashes etc. allow *restrained refinement* to be performed (118). The protein is still treated as a flexible entity, but penalties are given to movement of the structure too far from ideal values. This method of restrained refinement greatly reduces the number of parameters to be refined. At various stages of refinement, difference density maps are created, where sites wrongly filled with model show up as negative density and sites with density but no model show up as positive density. These difference maps are used as a guide for model correction. Another method of refinement called *simulated annealing* involves shaking up a structure and allowing it to slowly relax to a minimum (117). All the atoms are given a random displacement and allowed to relax. In cases where refinement is stuck with no obvious improvements to be made to the model, simulated annealing could be performed. At the final stages of refinement, water molecules are added into positive peaks in the difference density maps, for example those positive peaks that correspond to the 3σ contour level. This can be done manually or using features like the *add waters* feature in phenix.refine (121). At this stage, densities that are too big to be water molecules (>3σ) are carefully analyzed since other reagents from the crystallization condition or that purified along with the protein could have crystallized. In Rsr, one such site was identified with 6σ positive density and Calcium ion was refined into this site. Model building and refinement are iteratively done to obtain the structure that most accurately matches the observed diffraction data.

At every stage of refinement, a *Reliability factor (R Factor)* is calculated (117). The R factor is defined as the ratio of sum of differences between observed and calculated structure factor amplitudes to the sum of the observed structure factor amplitudes (117). A randomly selected set of reflections (for example, a random 5% of reflections) are kept aside as a test set and not used for structure refinement. After every step of refinement, using this test set, an R factor called the *Free R Factor (R_{free})* is calculated (117). Since R_{free} is calculated from data that has not undergone the refinement process, its measure is an unbiased indicator of the refinement process. Typically, for a well refined structure at a good resolution, the R factor is no more than 0.25 (or 25%) and the R_{free} is no more than 0.30 (30%).

$$R = (\Sigma_h \, ||F_{obs}|-|F_{calc}||) \, / \, (\Sigma_h \, |F_{obs}|)$$

$$R_{free} = (\Sigma_{test\ set} \, ||F_{obs}|-|F_{calc}||) \, / \, (\Sigma_{test\ set} \, |F_{obs}|)$$

The X-ray diffraction data sets were integrated, scaled and merged using HKL2000 (122). From the three wavelength MAD data from Se-met Rsr crystals, 13 out of 14 selenium sites were located using phenix.hyss (121). Initial phase calculation using data from 50 to 2.6Å resolution, followed by density modification was done using AutoSHARP (123). With these density modified phases, the electron density map was calculated and initial model was built using TEXTAL (120). Despite the low resolution of the MAD data and poor quality of the phases, TEXTAL built a nearly complete initial model. Improvements in the model were done manually on XtalView (124) using the structure of *X. laevis* Ro as reference (PDB ID 1YVR). This improved model was refined against the native Rsr data first by rigid body refinement and subsequently by simulated annealing refinement using phenix.refine (121). Although the Se-met derived crystal is sufficient for the entire process of phase calculation and structure solution, it is often the case that heavy atom derived crystals are extra-sensitive to radiation and do not produce the best possible quality of diffraction. In the case of Rsr, the Se-met crystals diffracted poorer than the native crystal, hence after phase calculation, the model was refined against the native, higher resolution data. AutoBUILD on the phenix suite was

used to improve the model. Bias minimized maps were created using the TB Bias Removal Server (125). Composite omit maps were created using CNS (126). Strong positive electron density was seen surrounded by the cation binding MIDAS motif (DxSxS…T…D) and Ca^{2+} was refined into this density since the crystallization condition contains 200mM Calcium acetate. The final model was refined with crystallographic R_{work} and R_{free} of 22.3 % and 26.2 % respectively. Refinement statistics are shown in **Table 2.3**. 99 % of residues are in the allowed region of the Ramachandran plot with 1% in the generously allowed region. Electrostatic surface potential was calculated using APBS (127) and UCSF Chimera; all other figures were prepared using UCSF Chimera (128),(129).

2.3.10 *in vitro* transcription of *D. radiodurans* Y RNA (DrY RNA)

The region corresponding to DrY RNA template was amplified from *D. radiodurans* genomic DNA (American Type Cell Culture) using primers shown in **Table 2.4** and ligated into the pETBlue1 vector. A blunt cutting restriction enzyme (Afe1 or HincII) recognition site was incorporated in the primer at the 3' end. Prior to *in vitro* transcription, the pETBlue1-DrY RNA[123] was digested with Afe1 or HincII. Analytical *in vitro* transcriptions were performed with varying concentrations of divalent ions (10 to 20mM), template concentration (10 to 15ng/µl) and dNTPs (4 to 8mM) in the reactions. The transcripts were analyzed on 5% denaturing polyacrylamide gels. The reaction with 10ng/µl DNA template, 6mM dNTPs, 15mM $MgCl_2$ produced the maximum amount of pure Y RNA of the desired size and was repeated on a large scale. Large scale transcription was performed at 40ºC for 9 hours following which the reaction was quenched using EDTA and Ethanol and frozen at -80ºC. After thawing, the sample was centrifuged at 8000rpm for 30 minutes. The pellet was dried with Nitrogen gas and resuspended in water for subsequent purification on a G-25 column. The eluate from the G-25 column was loaded on 6% preparative denaturing gels. The bands were visualized with UV shadowing, following which the required band was cut and the RNA was electro-eluted. As a final step of purification, a C-18 column was used. The eluate

from the C-18 column was heated to 72 °C for 5 minutes and allowed to fold at room temperature for 45 mins prior to use. To determine the homogeneity of the sample, the RNA was run on a non-denaturing polyacrylamide gel

2.3.11 Design of truncations of DrY RNA suitable for crystallization

Based on previous work on *X. laevis* Ro, where RNAse protection assays have helped identify the Ro binding region of Y RNA (89), (90) various truncations of DrY RNA were designed. In order to design truncations of DrY RNA which might adequately represent the secondary structure of the full length DrY RNA as well as bind Rsr, parts of the variable region of DrY RNA were deleted and the resulting truncated sequences were folded using mFOLD. Three truncations that folded identical to the full length DrY RNA were chosen for the study (**Figure 2.10**). These included a 35-mer fragment (DrY RNA^{35}), a 33-mer (DrY RNA^{33}) and a 39-mer with an engineered GAAA tetraloop (DrY RNA^{GAAA}). The RNA was resuspended in 25 mM Hepes pH 7.0, 10mM $MgCl_2$, 0.5 mM EDTA and 150mM NaCl and annealed by heating at 85 ºC for 5 minutes followed by cooling at room- temperature for 45 minutes. Folded RNA was subjected to non-denaturing gel electrophoresis on a 15% TBE Ready Gel (Biorad) to check for formation of alternate dimer conformations (**Figure 2.10**).

2.3.12 Fluorescence assays to study binding between Ro and DrY RNA^{35}

Intrinsic tryptophan fluorescence measurements were carried out on an ISS PC1 photon counting spectrofluorometer with Ro at a concentration of 50nM. The reaction buffer consisted of 25mM Hepes 7.0, 150mM NaCl, 10mM $MgCl_2$. With the concentration of Rsr equal to 50nM, RNA was titrated in to achieve final concentrations from 5.8nM to 665nM. Following each titration, the sample was allowed to equilibriate for 10 mins and fluorescence emission spectra (315 to 450 nm) were acquired upon excitation at 295 nm. Simultaneous titration of RNA to buffer alone was done and the tryptophan emission data was corrected for fluorescence from RNA or buffer components.

The effect of photobleaching on tryptophan fluorescence was measured by

repeatedly exciting the reaction buffer containing Rsr at 295nm and recording the fluorescence emission from 315 to 450 nm. The maximum photobleaching amounted to 5.8 % of the total fluorescence and was considered negligible.

To estimate the contribution from the inner-filter effect from RNA, RNA was titrated into the reaction buffer and the absorbance of RNA was measured at 295nm and 335nm on a Cary100 spectrophotometer. The fluorescence intensities were corrected for RNA inner-filter effect as:

$I_{corr} = I_{obs} * \text{antilog } [(A_{295} + A_{335}) / 2]$

Where I_{corr} is the corrected Intensity, I_{obs} is the observed Intensity, A_{295} and A_{335} are absorbance of RNA at 295 and 335nm respectively.

The fraction bound (f) was calculated by determining the decrease in corrected fluorescence intensity at 335nm of the sample compared to the decrease when Ro was completely bound by RNA. The data were fitted to a single site binding model, using the following equation to determine the value of the apparent binding constant (K_d).

$f = [RNA] / ([RNA] + K_d)$

Stoichiometry of binding was calculated as suggested in ref. (130)

2.3.13 Size exclusion chromatography of Rsr:DrY RNA[35]

Rsr was purified as described above, on a Ni^{2+} affinity column. Rsr at 3µM was mixed with a five fold molar excess of DrY RNA[35] and dialyzed against buffer containing 25mM Phosphate pH 7.0, 150mM NaCl and 2mM DTT. Following dialysis, the sample was concentrated by 10-fold and injected on a Superdex 200 analytical column equilibrated with the same buffer. All runs on the superdex 200 column were performed at a flow rate of 1ml/minute. Identical elution profiles were obtained with the complex at concentrations as low as 1µM. Different concentrations of protein alone (1µM to 500 µM) and DrY RNA[35] alone (2µM to 590µM) were injected in separate runs on the column. The column was calibrated with a mix of standard protein markers (Biorad # 151-1901). A standard curve of elution volumes versus Log_{10} of Molecular weights was plotted and used to calculate the molecular weights of the Rsr, DrY RNA[35] and Rsr:DrY

RNA complex peaks. Ratio of Absorbance at 260nm to absorbance at 280nm was measured for each fraction. Ratios greater than 1 were taken to be indicative of RNA. To detect protein in the fractions, the fractions were spotted on a nitrocellulose membrane, incubated successively with Anti-6his antibodies and Anti-mouse IgG Alkaline phosphatase and stained with SigmaFAST^{TM} BCIP/NBT.

2.3.14 Electron microscopy on Rsr:DrY RNA^{35} complex (by Christos G. Savva and Andreas Holzenburg)

Rsr:Dr YRNA^{35} (~3µM in 25mM Phosphate pH 7.0, 150mM NaCl, 2mM DTT) samples were prepared for electron microscopy according to Valentine (131) and stained with 2%w/v aqueous uranyl acetate pH 4.3. Duplicate grids were prepared similarly but stained with 2%w/v ammonium molybdate pH 7.0. Samples were examined using a JEOL 1200EX transmission electron microscope operating at 100 kV. Micrographs were recorded at a calibrated magnification of 38,030. Selected micrographs were digitized using a Leafscan 45 microdensitometer at a 20-µm scan step corresponding to 5.26 Å/pixel at the specimen level. Particles were manually selected using the BOXER program from the EMAN (132) single particle analysis software package. Approximately 1600 particles were chosen and further processed with IMAGIC-5 (133). Particle images were normalized, band-pass filtered and subjected to reference free classification. Non-redundant class averages were then chosen as references for multireference alignment (134). The process was iterated until stable classes were obtained based on visual inspection and intraclass variance. All class averages are displayed without imposing symmetry.

2.3.15 Crystallization of Rsr:DrY RNA complex

Prior to crystal trials, sample was prepared by mixing Rsr and DrY RNA^{35} at equal molar ratios (~490µM each) and incubating for 1 hour on ice. The sample was then given a 10 minute spin at 15000Xg and used for crystal trials. Initial trials were performed using screens Crystal Screen 1 and 2 and Natrix purchased from Hampton

Research Inc. and Wizard Screen 1 and 2 from Emerald Biosciences. In these screens the primary variables are salt, pH and precipitant. Various conditions with different precipitants yielded crystals, for example CS1#26. Optimization of crystals was done by a Grid method which consists of systematically varying pH and concentrations of salt and precipitant around the initial condition which yielded crystals. Additive screens and Detergent screens purchased from Hampton Research Inc., were also used to optimize the crystals. These screens comprise of a variety of small molecules and detergents that affect crystallization by manipulating sample-solvent and sample-sample interactions. These crystals were highly reproducible and grew even when Se-met substituted Rsr was used in place of native Rsr. The best crystals after optimization, were obtained in Crystal Screen 1#26 (0.2M ammonium acetate, 0.1M Sodium citrate pH5.6 and 30% MPD) in hanging drop plates. The crystals were frozen in mother-liquor as the cryoprotectant because of the presence of 30%MPD in the crystallization solution. The best crystal diffracted to 6 Å at beamline 23-ID, with P6 symmetry. Detector distance during data collection was around 700mm for optimal spot separation, considering the unusually large cell dimensions. Data statistics are compiled in **Table 2.1**.

Crystals were also obtained by co-crystallization of Y RNA and Rsr post Thrombin cleavage. These crystals looked identical in shape and size to the crystals grown with 6his-Rsr:Y RNA, but the diffraction quality did not improve with crystal optimization.

2.3.16 Structure determination of Rsr:DrY RNA complex

Although low resolution, the diffraction data obtained from crystals of the Rsr:DrY RNA complex was used for Molecular replacement, using the structure of Rsr as a search model was performed using Molrep, Amore and Phaser programs on the ccp4 suite. 12 to 14 molecules were searched for in a single asymmetric unit.

2.4 ADDITIONAL MATERIAL

2.4.1 Identifying Y RNAs in *M. smegmatis* and mycobacteriophage Bxz1

In *M. smegmatis* and the mycobacteriophage Bxz1, sequences for Ro orthologs have been identified in the genome, though there are no reports of the presence of Y RNAs. We have developed an *in silico* method to identify potential Y RNA sequences in these genomes, in collaboration with Dr.Thomas R. Ioerger, Department of Computer Science, Texas A& M University. The approach is to write a program which can scan a genome sequence given as input, for the presence of sequences following the criteria observed in known Y RNAs like size restrictions, presence of conserved residues in the Ro binding box and complementarity between the 5' and 3' ends. The hits from the program are then folded on mFOLD to select those sequences which when folded, resemble the predicted structures of known Y RNAs. The chosen sequences are candidate Y RNAs.

In collaboration with Dr. Thomas R. Ioerger, a program was written, which looks for sequences in genomes, which are 60 to 120 nucleotides long, have complementary 5' and 3' ends and have the conserved sequence of base pairs present in the Ro binding helix of known Y RNAs. Using the *D. radiodurans* genome sequence as a test for this program, 11 hits were obtained. One of them was the sequence which has previously been identified as a dY RNA (4). This was considered a positive test for the program**.** The *M. smegmatis* genome showed 13 hits and the Bxz1 genome showed 0 hits. Upon folding the MsY RNA hits using mFOLD, at least 3 of them resembled Y RNAs (**Figure 2.21**). These candidate Y RNAs will be used for binding studies with MsRo and BzRo.

2.4.2 Studies on Ro from the mycobacteriophage Bxz1

BzRo with an N-terminal 6his tag followed by an MBP tag was purified on a Ni^{2+} chelating sepharose column (**Figure 2.22**). The eluate from the Ni^{2+} column consisted of >95% pure fusion protein as judged by SDS-PAGE. The fusion protein at a concentration of 1mg/ml was cleaved with TEV Protease at a ratio of 1:100 (protease: fusion-protein) during dialysis for 16 hrs in 4 ºC, against buffer containing 50mM Hepes

pH7.0, 500mM NaCl and 2mM DTT. The TEV Protease used, had an N-terminal 6his tag and was overexpressed using a construct available in the lab. The protease was

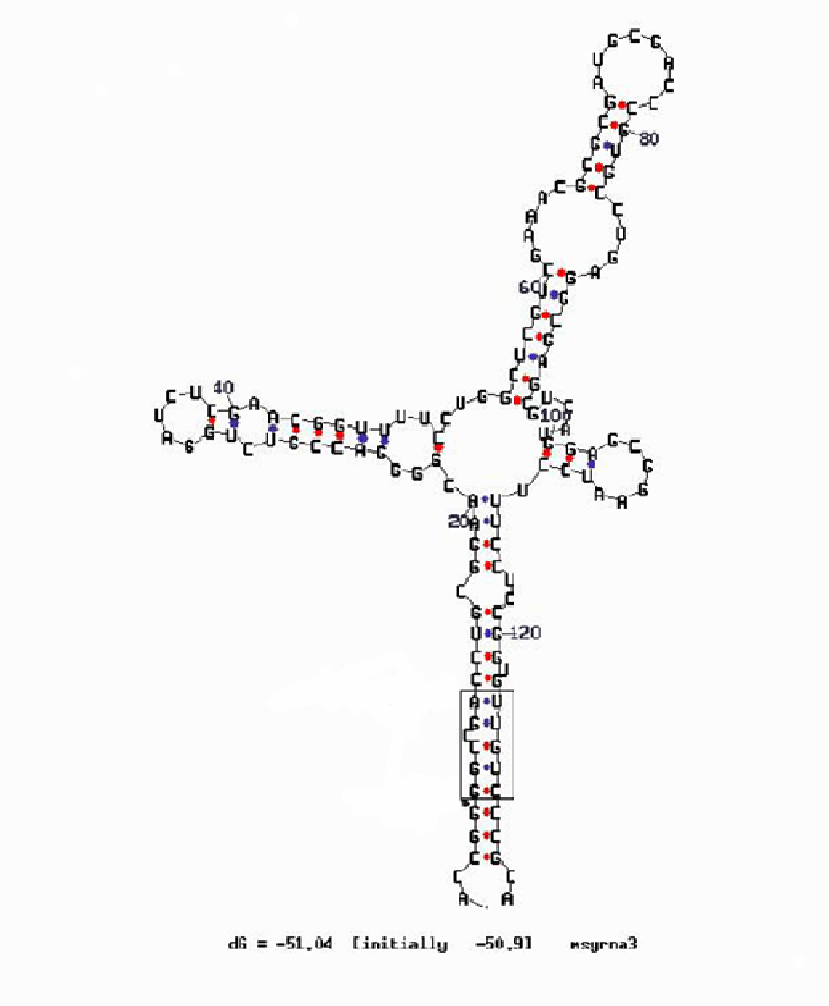


Figure 2.21 mFOLD secondary structures of a candidate Y RNA in the *M. smegmatis* genome. Putative Ro binding site is boxed.

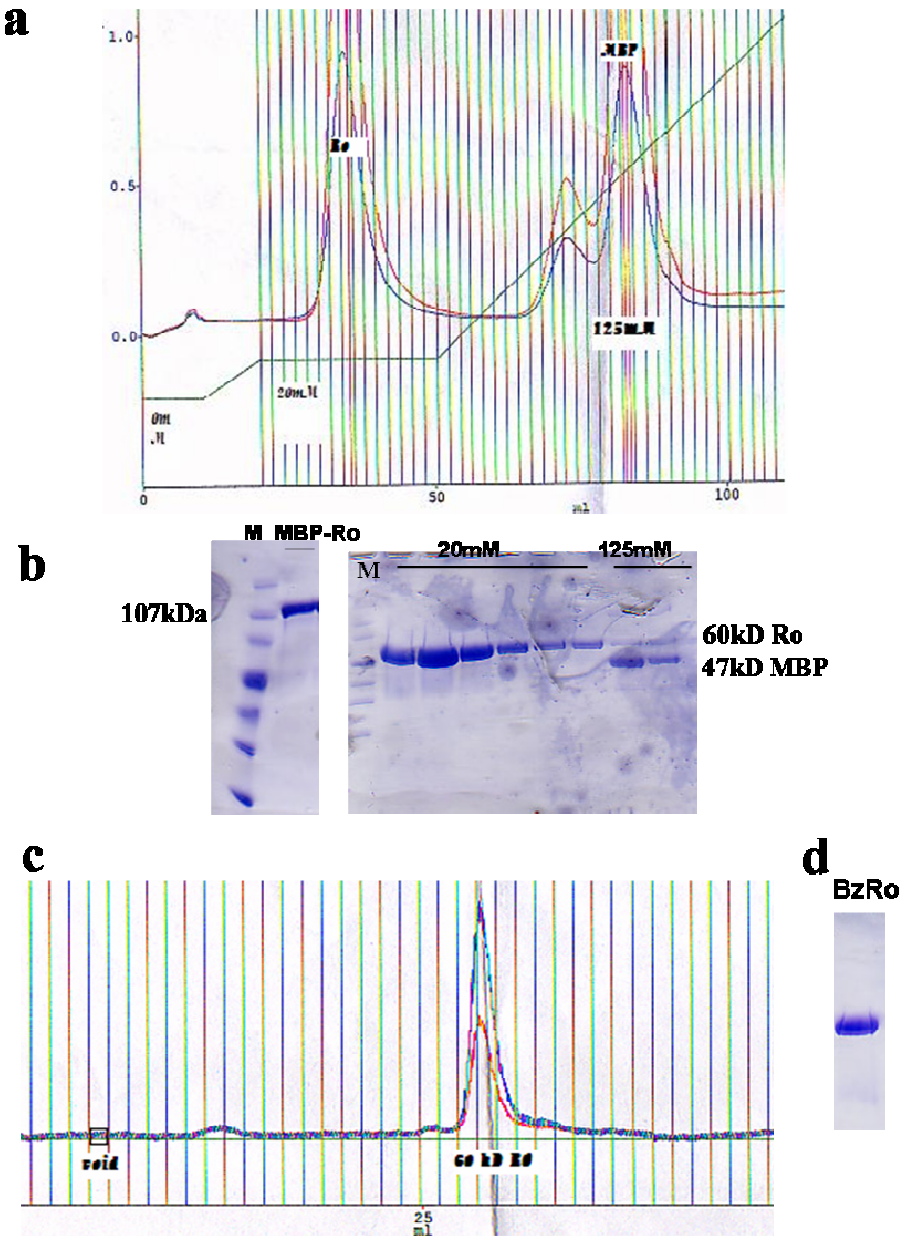


Figure 2.22 Purification of BzRo

a. Plot of absorbance at 280nm versus volume shows the elution profile of BzRo from a Nickel column post TEV cleavage. Steps corresponding to increasing concentrations of Imidazole (green) are shown. Pure cleaved BzRo elutes at 20mM Imidazole. 6his-MBP elutes at 125mM Imidazole. b. SDS-PAGE analysis of the fractions shows the fusion protein, cleaved BzRo and 6his-MBP. c. Elution profile from an S-200 column shows that freshly purified BzRo is homogeneous. d. SDS-PAGE analysis of cleaved BzRo sample used for crystallization shows >90% pure protein.

purified on a Ni^{2+} column followed by gel-filtration on a superdex-75 column. The cleaved fusion protein was passed again on a Ni^{2+} column. BzRo eluted with 20mM Imidazole, whereas 6his-MBP eluted with 125mM and 6his-TEVProtease with 400mM Imidazole, thereby allowing us to separate Ro from the MBP and TEV protease (**Figure 2.22**). Pure and homogeneous BzRo was concentrated to 3mg/ml but was unstable in solution and showed marked precipitation after 3 days. This phenomenon posed a hinderance to crystallization, though the protein could still be used for Y RNA binding assays. Led by the results of the Rsr stabilization in phosphate buffer, BzRo was lysed and purified in phosphate buffer, but BzRo tended to be unstable and precipitated even in this buffer.

2.4.3 Suggested reading

"Crystal structure of Rsr, an ortholog of the antigenic Ro protein, links conformational flexibility to RNA binding activity." By Ramesh et al., *Journal of Biological Chemistry,* March 2007, is available at DOI 10.1074/jbc.M611163200.

CHAPTER III
STRUCTURAL STUDY OF CsoR

3.1 INTRODUCTION

3.1.1 Role of copper in living organisms

Copper is an essential element required in trace amounts in living cells (135). Being a redox-active element, copper is utilized by many enzymes that catalyze reactions involving oxidation and reduction. In biological systems, copper mostly exists in two oxidation states- Cu(I) and Cu(II). Despite their different sizes and stereochemistries, both Cu(I) and Cu(II) have an equal tendency to form complexes with organic ligands (135). Hence, they make a convenient redox-couple that is useful in the range of redox potentials that are encountered in biological systems (135). Unlike ions like Na^+ and K^+, which are very mobile in biological media and form complexes with relatively low stability, $Cu^{+/++}$ ions are static in biological media and form complexes of much higher stability (135). The preferred donor atoms for metal ions are nitrogen, sulfur and oxygen making histidine, cysteine/ methionine, aspartic acid/glutamic acid the typical ligands present in metal sites of proteins (135). Of these, Nitrogen and Sulfur are the preferred ligands for Cu(I).

Copper utilizing enzymes have thus evolved to exploit the Cu^+/Cu^{2+} redox-couple. Cytochrome *c* oxidase, a copper utilizing enzyme and the terminal enzyme in the respiratory chain, is present in the inner membrane of mitochondria and bacteria (136), (137). Cytochrome *c* oxidase catalyzes the reduction of dioxygen to water, and contains two heme and two copper sites, CuA and CuB. While the CuA site primarily acts as an electron transfer centre, the CuB site is involved in stabilizing an intermediate peroxide bridge which subsequently gets reduced to form water (138). Cu/Zn Superoxide dismutase (SOD) is another enzyme that uses the reduction and oxidation of copper. Cu/Zn SOD dismutates superoxide to oxygen and hydrogen peroxide. Superoxide radicals cause oxidative damage in the cell by targeting proteins (139).

Hence, by the dismutation of superoxide, SOD acts as a key antioxidant in the cell (140). Copper deficiency in cells results in diminished SOD activity and this results in cellular oxidative damage. Yet another class of copper-dependent enzymes are the amine oxidases that catalyze the oxidation of biologically important mono and di amines in cells (141). Copper amine oxidases (CAO) are widely dispersed across species and are present in bacteria, fungi, plants and mammals (142). In bacteria and fungi, CAOs are involved in the metabolism of amines which act as the primary source of carbon and nitrogen for growth (142). Copper-dependent enzymes include Laccases, Tyrosinases, Haemocyanin (involved in transport of O_2), Lysine oxidases (Cross linking of collagen), Ascorbate oxidase, Galactose oxidase and the Blue proteins (involved in electron transfer) (135). Despite its varied uses in fundamental metabolic processes in the cell, uncomplexed copper is highly toxic to the cell.

3.1.2 Damage and toxicity by Cu

Cu(I) is highly unstable and gets readily oxidized to Cu(II), causing reduction of molecular oxygen to form hydrogen peroxide. Reduction of Cu(II) back to Cu(I) by cellular reductants like ascorbate and glutathione results in the generation of reactive Oxygen species (ROS) (143). In the presence of superoxide or reducing agents like ascorbate, Cu(II) undergoes reduction to Cu(I) accompanied by the conversion of hydrogen peroxide to hydroxyl ($^{\cdot}OH$) free radicals, one of the most oxidizing free radicals encountered in biological systems (143), (144). Transition metals like copper, that are more prone to oxidation, have a higher ability to generate harmful free radicals. ROS like the hydroxyl radical ($^{\cdot}OH$) and superoxide ($^{*}O_2^{-}$) cause oxidation of protein side-chains, protein-protein cross links and oxidation of the protein backbone (145,146).

DNA strand-breaks and oxidation of nucleic acid bases are also a consequence of copper mediated free radical damage in cells (147), (148), (149). An excess of copper in the cell is also known to cause peroxidation of membrane lipids (150).

Even when bound to protein, copper can react with hydrogen peroxide to form hydroxyl free radicals ($^{\cdot}OH$) (151). These free radicals oxidize neighboring amino acid

residues in the protein. This type of metal catalyzed oxidation occurs at sites on the protein that bind metal ions. Metal catalyzed oxidation has been implicated in many neurodegenerative diseases like Alzeimers (152), Parkinsons (153), prion diseases (151) and aging (154). The Cu(II) binding prion protein PrP has been shown to switch from its soluble cellular form (PrP^{c}) to a scrapie form (PrP^{sc}) and this switch has been shown to be critical for prion infectivity (151). Reduction of Cu(II) by the reductant ascorbate in the presence of oxygen results in oxidation of PrP and subsequent aggregation to a PrP^{sc} like form (151). During dialysis related amyloidosis, fiber formation by the β-2-microglobulin polypeptide has been shown to be catalyzed by Cu(II) (155). Formation of intermediate native-like conformations due to Cu(II) catalyzed backbone isomerization and stabilization of amyloid-like conformations by Cu(II) are critical steps in fiber formation (155).

Considering that copper is an essential trace element for numerous biological reactions but is responsible for oxidative damage and cellular toxicity, cells must maintain a balance in the amount of copper present in them. This process of maintaining the optimal amount of cellular copper is called copper homeostasis.

3.1.3 Copper homeostasis mechanisms

In general, maintenance of metal concentrations in the cell occurs in three steps: uptake, export, sequestration and regulation.

Copper uptake: Uptake of copper in bacteria and yeast, is performed by permeases or import proteins and is controlled either by regulating the import proteins or by limiting the amount of metal available for uptake (156). Integral membrane permeases or import pumps may act in a passive energy-independent manner or may utilize energy from ATP hydrolysis to facilitate copper uptake (157), (158), (159). In mammals, copper uptake is mediated by the plasma protein ceruloplasmin or in the free form through a membrane bound copper reductase (160). Both pathways work in an energy independent manner with copper being transported into the cell without the intake of the carrier protein

ceruloplasmin or other carrier ligands (160).

Copper export: Export of copper from cells is done by primary and secondary active transporters. While primary active transporters use energy from the hydrolysis of ATP, the secondary active transporters couple the export of copper with the membrane potential (156). The most common mechanism of copper export involves the P-type ATPases. The cation transporting P-type ATPases are a superfamily of primary active transporters (161). These are membrane-residing, ATP hydrolyzing pumps that are highly conserved in sequence from bacteria to mammals. In general, the pump consists of an N-terminal phosphorylation domain, an ATP binding site, a phosphatase domain and the C-terminal ion translocating channel (161). P-type ATPases undergo phosphorylation at a conserved Aspartic acid residue present in a seven residue signature motif of D-K-T-G-T-[L I V M]-[T I S] found in all P-type ATPases. Phosphorylation is accompanied by a conformational change that allows ion translocation. A model based on the Menke's ATPase from humans suggests that the N-terminal regulatory domain binds copper at each of its six M-X-C-X-X-C motifs. Binding of copper to the regulatory domain allows the phosphorylation site to be accessible and this allows the subsequent steps in translocation to take place (156). Mutations in the Menke's ATPase and its homologous Wilson's ATPase in humans are associated with the copper excretory disorders called Menke's and Wilson's diseases (162). The copper-dependent enzymes show diminished activity in these disorders. Although the general copper binding N-terminal fold is conserved among the cation transporting ATPases, the number of units composing the N-terminal could differ. For example, while the Menke's protein has 6 metal binding sites, the CtaA and PacS ATPases from the cyanobacterium *Synechocystis* has only one such fold (163), (164).

Copper sequestration: Sequestration of copper is important since free copper can be toxic to the cell. There are various mechanisms used for copper sequestration. One of the primary protein components involved in the sequestration of Cu(I) is the

metallothionein. Metallothioneins are cysteine-rich, 25 to 60kDa proteins that form polynuclear Cu-thiolate clusters (165). Cysteines in metallothioneins are arranged in repeats of the Cys-Cys, Cys-X-Cys or Cys-X-X-Cys motifs. Metallothioneins typically have few hydrophobic residues and the core of the protein is stabilized by metal-thiolates. The number and type of metal bound by the metallothionein often determines the fold adopted by the protein.

Another mode of Cu(I)-sequestration seen in plants, fission yeast and some microorganisms involves the phytochelatins. These are γ-glutamic acid- cysteine polymers derived from glutathione, with a general structure of $(\gamma\text{-Glu-Cys})_n\text{-Gly}$ (166). Phytochelatins are involved in metal detoxification *in vivo* and possibly in reactivation of the copper-dependent enzyme diamino oxidase as seen *in vitro* (167).

Copper chaperones represent an important class of proteins involved in copper sequestration. Copper chaperones deliver copper to copper utilizing enzymes or to the copper export machinery (168). In eukaryotes, copper chaperones deliver copper to the trans-Golgi network, the Cu/Zn SOD in the cytoplasm and help in the assembly of cytochrome *c* oxidase in the mitochondria (169), (136). In yeast, the chaperone CCS mediates insertion of copper into the Cu/Zn SOD presumably in copper limiting conditions in the cell (169). In prokaryotes, chaperones like Atx1 and CopZ (discussed below) deliver copper to specific membrane bound copper export proteins (170). The CCS chaperones carry the Cys-X-X-Cys motif similar to SOD and the Atx1-like copper chaperones carry the Met-X-Cys-X-X-Cys motif similar to that found in the regulatory N-terminal domains of P-type ATPases (171), (168). Structural and biochemical characterization of a number of copper chaperones has shown that these proteins possess similar metal binding folds. Studies on the Atx1 chaperone suggest that copper is coordinated in an exposed loop region which is flexible enough to allow changes in coordination number and geometry (172), (173). These changes probably occur when the chaperone interacts with the copper-dependent enzyme or transporter protein.

Regulation: Regulation of the different proteins involved in uptake, export and

sequestration of copper is essential for maintenance of the optimal concentration of copper in cells. In eukaryotes, regulation of these proteins takes place both transcriptionally and post-translationally. In prokaryotic organisms however, majority of the regulation happens at the level of transcription (174). Proteins involved in transcriptional regulation respond to copper by directly binding copper and causing activation, repression or derepression of one or more copper utilizing enzymes or copper homeostasis proteins. The key factor in regulation is the recognition of the correct metal by the regulatory protein. This specificity for the metal is achieved by virtue of the distinct coordination chemistries and geometry presented at the metal binding sites (174) since each metal has a preference for certain coordination geometry and coordinating ligand types. Based on structural similarities, prokaryotic transcriptional regulators have been classified into six major families (174), (175). These are the MerR, ArsR/SmtB, Fur, DtxR and NikR and the newly discovered CopY families. Regulatory proteins within a family may respond to different metals (174). Members of the MerR family repress the operon until they recognize specific metals. Upon metal binding, a conformational change induced in the operator region allows RNA polymerase to initiate transcription. Recently, in *E. coli,* a Cu(I) responsive regulator CueR from the MerR family has been characterized (14). CueR regulates the expression of CopA (a P-type ATPase) and CueO (a multicopper oxidase). Cu(I) specificity of CueR is likely due to the special coordination environment created at its metal binding site. CueR binds Cu(I) in a linear two coordinate manner, involving two conserved Cysteine residues (176). The ArsR/SmtB family of regulators consists of proteins that are bound to the promoter DNA region in their apo state, thereby repressing the operon. Metal binding results in allosteric regulation of DNA binding causing derepression of the operon (177). A Cu(I)/Ag(I)/Zn(II)/Cd(II) responsive regulator belonging to this family is the BxmR protein from the cyanobacterium *Oscillatoria brevis*. BxmR regulates the expression of BmtA (a metallothionein) and Bxa1 (a CPx-ATPase) (178). Most of our understanding of this family comes from work on the arsenic responsive ArsR protein from *E. coli* and the zinc responsive SmtB protein (179), (180). Members of the DtxR, Fur and NikR

families possess weak affinity for DNA in their apo state. In the metal bound state, they bind DNA to repress the operon. For example, IdeR an iron responsive regulator from *M. tb* from the DtxR family regulates iron uptake genes (181). Upon sensing iron in the cytoplasm, IdeR binds DNA to repress iron uptake proteins. Fur family proteins regulate not only proteins required for iron acquisition but also enzymes involved in the synthesis of iron sequestering siderophores (182). The NikR family so far consists of only nickel responsive regulators (174). Copper responsive regulators belonging to the DtxR, Fur and NikR families are yet to be identified. CopY is a copper responsive transcriptional repressor of copper ATPases CopA and CopB from *Enterococcus hirae* (183).

Regulators from all these families bind DNA as oligomers- either dimers or tetramers, with the metal coordinated in a subunit-bridging site (174). Another feature common to many of these families is the presence of a winged helix domain or a modified helix-turn-helix motif (174). Currently no structure is available for a MerR, ArsR, Fur, CopY family metalloregulatory protein in complex with DNA. Biochemical characterization and structural similarity with other proteins suggests that these motifs are involved in DNA binding. High resolution DNA bound structures are available for DtxR and NikR family members Ni(II)-DtxR (184) and Ni(II)-IdeR (15) and the Ni(II)-NikR (23). Mechanisms of copper homeostasis differ in eukaryotic versus prokaryotic organisms.

3.1.4 Copper homeostasis in yeast and mammals

Various Cu resistance mechanisms have been studied in yeast. In *Saccharomyces cerevisiae* sequestration of Cu(I) is done by metallothioneins. Expression of the *S. cerevisiae* metallothionein (ScMT) is under the control of a Cu(I) sensing transcriptional activator (Ace1). Ace1 binds Cu(I) to form a polynuclear Cu(I)-thiolate cluster (185). Upon binding Cu(I), Ace1 undergoes a conformational change that increases its affinity for the Upstream Activation Sequence (UAS), thereby increasing transcription of ScMT. Although ScMT can bind other metals, its upregulation occurs only in the presence of Cu(I). Another transcriptional factor, Mac1 binds Cu(I) to form a polycopper cluster

similar to that formed by Ace1. Mac1 regulates transcription of at least six genes, two of which encode Cu-permeases Ctr1 and Ctr3 and one encoding the metalloreductase Fre1. In conditions with limiting Cu, Mac1 activates transcription of these genes. In the presence of excess Cu(I), Mac1 activation of these genes is inhibited, thus controlling the level of Cu(I) in the cell (186). Mac1 possesses a Cu(I) binding motif –C-X-C-X_3-C-X-C-X_2-C-X_2-H. A similar motif is found in the Cuf1 transcriptional factor in *S. pombe.* In *S. pombe,* Cu(I) sequestration is also done using phytochelatins (187). In *Candida glabrata* although phytochelatins are used for sequestration of other metals like cadmium, copper sequestration is done primarily by metallothioneins. Copper export pumps have not been identified in *S. pombe* and *C. glabrata.*

In mammals, there exist redundant mechanisms for Cu(I) homeostasis. Mammals have both metallothioneins to sequester Cu(I) as well as ATPase pumps for export of Cu(I). Metallothioneins from mammals and yeast share little similarity besides the Cu(I) binding motif. In contrast, Cu(I) export ATPases are highly conserved from bacteria to mammals (188). Between the Cu(I) exporting ATPases and metallothioneins, common Cu(I) binding motifs have been identified like the M-X-C-X-X-C motif in the Menkes ATPase (MNK) from humans. Post translational control of Cu(I) uptake protein Ctr1 during conditions with high cellular Cu(I) ions has been shown (189). Endocytosis of the plasma membrane results in decrease in the concentration of Ctr1 in the membrane, reducing the uptake of Cu(I).

3.1.5 Copper homeostasis mechanisms in prokaryotes

Bacterial cells maintain close to micromolar levels of copper (190). However, most of the copper utilizing enzymes are present in the periplasm or attached to the inner or outer membranes and there is no evidence yet for copper utilizing enzymes in the cytoplasm of bacterial cells. In *E. coli,* a gram negative bacterium, copper utilizing proteins like amine oxidase, superoxide dismutase and cytochrome oxidase are present in the periplasm. In the gram positive bacterium *B. subtilis,* in the absence of a periplasmic compartment, the cytochrome oxidase is located on the plasma membrane and a copper

utilizing laccase is present in spore coats, outside the cytoplasm (191). In many cyanobacteria, membrane bound organelles called thylakoids are present in the cytoplasm. In these organisms, electron transfer during respiration and photosynthesis occurs in the thylakoids, with the help of copper proteins like plastocyanin and cytochrome *c* oxidase (192). Even in these proteins, the copper sites face the thylakoid lumen and not the cytoplasm. Since there is no obvious requirement for copper in the cytoplasm of bacterial cells, and unbound Cu(I) in the cytoplasm is highly toxic to the cell, bacterial homeostasis mechanisms must be directed towards transport of Cu(I) to the periplasm or the plasma membrane and out of the cell. In the last couple of decades, numerous proteins have been identified in the cytoplasm, periplasm and in the membranes, that play a role in copper homeostasis in bacterial cells.

In *E. coli,* the CusCFBA system has been identified, controlled by the CusR/S system. CusCBA together form a membrane bound copper export system that traverses the periplasm. CusF is a periplasmic chaperone protein which directs excess Cu(I) in the periplasm towards export via CusCBA. The *cus* operon itself is regulated by the two component system CusR/S which senses excess copper in the periplasm. Another mode of copper homeostasis in *E. coli* consists of the CueR/CueO/CopA system. CopA is an inner membrane P-type ATPase that exports Cu(I) out of the cytoplasm. CueO is a multi-copper oxidase that binds four copper ions and gets transported from the cytoplasm to the periplasm via the Twin Arginine Transporter (TAT) pathway. Expression of both CopA and CueO is regulated by a cytoplasmic Cu(I) responsive transcriptional regulator called CueR.

In *B. subtilis* and *Enterococcus hirae,* CopA functions as a P-type ATPase transporter of copper while CopZ, a cytoplasmic protein functions as a copper chaperone. Both CopA and CopZ are regulated by copper sensing transcriptional regulators viz. CueR in *B. subtilis* and CopY in *E. hirae* (14), (175).

In the cyanobacterium *Synechocystis* PCC6803, the presence of copper proteins plastocyanin and cytochrome *c* oxidase in the thylaoids within the cytoplasm requires copper to be trafficked though the cytoplasm. An elegant system has evolved to achieve

this trafficking using copper chaperones and copper transporters. In this system, copper is exchanged between transporters and the chaperone without release of free copper in the cytoplasm. Two P-type ATPases CtaA and PacS have been identified in the inner menbrane and the thylakoid membrane respectively (163), (164). CtaA imports copper into the cytoplasm but release of copper presumably occurs only upon interaction with the cytoplasmic copper chaperone Atx1. Atx1 then delivers the copper to the PacS transporter on the thylakoids membrane and once again release of copper is facilitated by the interaction between the two proteins.

Copper along with iron and zinc plays an especially important role in intracellular pathogens. Proteins involved in metal homeostasis or metal acquisition are known to be virulence factors themselves or regulate expression of virulence factors (193), (194), (195), (196). One such pathogen in which metal-binding is closely associated with virulence and pathogenesis is the tuberculosis causing bacterium-*Mycobacterium tuberculosis.*

3.1.6 Role of metals in *Mycobacterium tuberculosis*

One of the most remarkable features of *M. tuberculosis* which makes it such a successful pathogen is it's ability to survive within the phagosomal compartment in the host. *M. tb* has various mechanisms to counter the oxidative burst in the macrophage upon infection of the host (197). Especially iron, copper and zinc are involved in the oxidative stress response in the macrophage. Superoxide dismutases SodA and SodC in *M. tb* contain Fe and a Cu/Zn pair respectively. SodA is presumed to be involved in responses to endogenous oxidative phosphorylation. SodC is located in the periplasm and is involved in protection against host generated free radicals in activated macrophages (198). To control the level of intracellular metal ions, *M. tb* uses a variety of transporters including at least twelve P-type ATPases. Sequence based comparisons suggest that *M. tb* has close to 28 metal transporters, amounting to 24% of all the known transporters in *M. tb (197).* Of the twelve P-type ATPases, seven (ctp A, B, C, D, G, J and V) resemble 'typeIB' ATPases which are heavy-metal pumps involved in the transport of Cd^{2+},

Zn^{2+}, $Cu^{+/2+,}$ Ag^{+}, Mn^{2+}. CtpA, CtpB and CtpV are predicted to be $Cu^{+/2+}$ transporters. While CtpA and CtpB clearly possess an N-terminal metal binding motif of paired Cysteine residues characteristic of typeIB pumps, CtpV does not possess any recognizable motifs. ORFs upstream of CtpV contain histidine-rich metal binding motifs, suggesting a role for CtpV in metal transport. Besides transporters and metal utilizing enzymes, *M. tb* also possesses siderophores called mycobactins, a class of soluble low molecular weight molecules associated with the cell envelope (199). Mycobactins in *M. tb* have a high affinity for Fe^{3+} ions and have been shown to play a role in its survival in the macrophage.

Regulation of these key enzymes, transporters and mycobactins is of prime importance during survival of *M. tb* in macrophages. This regulation is done primarily at the level of transcription, by metal responsive transcriptional regulators. At least three such regulators have been identified in *M. tb* – FurA, IdeR and SirR. FurA, belongs to the Fur family of transcription regulators which responds to the changing amounts of iron and regulates expression of the downstream gene encoding KatG (a catalase peroxidase) and other virulence factors (200). IdeR, a member of the DtxR family of transcriptional regulators also responds to iron levels and is likely to regulate the expression of close to 40 genes including the *mbtA, mbtB, mbtI* gene cluster that encodes mycobactins (201). Other metal responsive regulators have been characterized in *M. tb,* like the Pb(II)/Cd(II) responsive CmtR (202) and the Ni responsive NmtR (203). The role of Cu as a cofactor in key enzymes like the Cu/Zn SodC and the presence of a number of Cu specific transporters in *M. tb* suggests a requirement for Cu specific regulation in *M. tb.* Yet, little is known about the mechanisms of Cu responsive regulation in *M. tb.* As part of the work described in this chapter, a copper responsive operon named *cso* (copper sensitive operon) has been identified in *M. tb* (**3.4.1**).

3.1.7 The Cu(I) responsive *cso* operon of *M. tb*

The *cso* operon (**3.4.1**) consists of three genes- *rv0967*, *rv0968* and *rv0969*. These three genes are transcribed as a single polycistronic message. Both *rv0967 and rv0969* are

upregulated in the presence of Cu(I) ions. *rv0969* has been annotated as *ctpV* and encodes a P-type ATPase that is predicted to be involved in copper transport (161). Although there have been no studies so far that suggest a role for the protein encoded by *rv0968*, sequence based analysis suggests that the N-terminal region of the *rv0968* gene product forms a 15-20 amino-acid transmembrane helix (http://www.doe-mbi.ucla.edu/TB/). The C-terminal region of this protein is predicted to be unstructured and possesses two histidine residues and many glutamate residues. Based on the sequence analysis alone, we speculate that this protein could be involved in Cu(I) exchange of some kind, possibly facilitating delivery of Cu(I) to the P-type ATPase encoded by *rv0969*. Our work shows that the 119 amino-acid protein encoded by *rv0967* is a Cu(I) sensing transcriptional regulator of the *cso* operon (**3.4.1**) and consequently, we have named this protein CsoR. Out of many metals, copper is the most potent inducer of the *cso* operon. CsoR specifically binds Cu(I) with 1:1 stoichiometry and an affinity higher than 10^{19} M^{-1}. Besides directly binding Cu(I), CsoR also directly interacts with DNA. Our results show that CsoR specifically binds DNA upstream of the *cso* operon. The upstream sequence bound by CsoR resembles the consensus *M. tb* promoter, with a ribosomal binding site, the -35 and -10 promoter elements and a transcription initiation site. As observed in most prokaryotic operator/promoter regions, this segment of DNA is pseudopalindromic. CsoR binds DNA only in the absence of Cu(I) and Cu(I) causes dissociation of preformed CsoR:DNA complexes. Put together our results show that Cu(I) negatively regulates the binding of CsoR to the operator/promoter region of the *cso* operon. This results in the derepression of the *cso* operon.

3.1.8 Significance of this work

Prior to our work, very little was known of the mechanistic details of CsoR regulation of the *cso* operon. In this chapter, we describe the structural characterization of CsoR. Using X-ray crystallography we have solved a high-resolution structure of Cu(I) bound CsoR. The structure not only helps us understand the basis of Cu(I) specificity and the

structural requirements for Cu(I) binding, but also provides valuable clues to the mechanism of allosteric regulation between Cu(I) binding and DNA binding by CsoR .In addition, this work contributes to our overall understanding of copper homeostasis in *Mycobacterium tuberculosis.*

3.2 RESULTS AND DISCUSSION

3.2.1 The overall crystallographic structure of Cu(I) bound CsoR

As described in **section 3.3.2**, different forms of CsoR were crystallized. Crystals of Apo:CsoR and Cu(I):CsoR are shown in **Figure 3.1**. The quality of diffraction from native and se-met Apo:CsoR crystals was not good enough for structure determination hence the se-met data was used only for phase calculation. The structure of Cu(I) bound CsoR was solved using diffraction data from Cu(I):CsoR crystals.

As described in Liu et al., "The structure was determined at a resolution of 2.55 Å, and the final model was refined to an R_{work} of 23.05% and an R_{free} of 27.87% with good stereochemistry. The refined model comprises $CsoR_{4–88}$, one Cu(I) atom and 20 molecules of water. Clear electron density was not visible for $CsoR_{89–119}$, which were either disordered in the crystals or lost to proteolysis during crystallization. These residues correspond to the C-terminal tail of Mtb CsoR, which is absent in other DUF156 homologs. The loop comprising $CsoR_{71–74}$ also showed weakly defined side chain density. Though the crystallographic asymmetric unit consists of one protomer of CsoR, applying crystallographic symmetry generated a homodimer (Fig. 5a), which buried 3,791 $Å^2$ (29.6%) of the accessible surface area between the two protomers. Given the considerably large buried surface area, the participation of conserved residues in the dimer interface, and the analytical equilibrium ultracentrifugation experiments that suggest a predominantly dimeric assembly state for both apo-CsoR and copper bound CsoR in solution, we propose that this crystallographic dimer is the physiologically relevant unit of CsoR. A dimer of dimers observed in the crystal buries only 2,407 $Å^2$

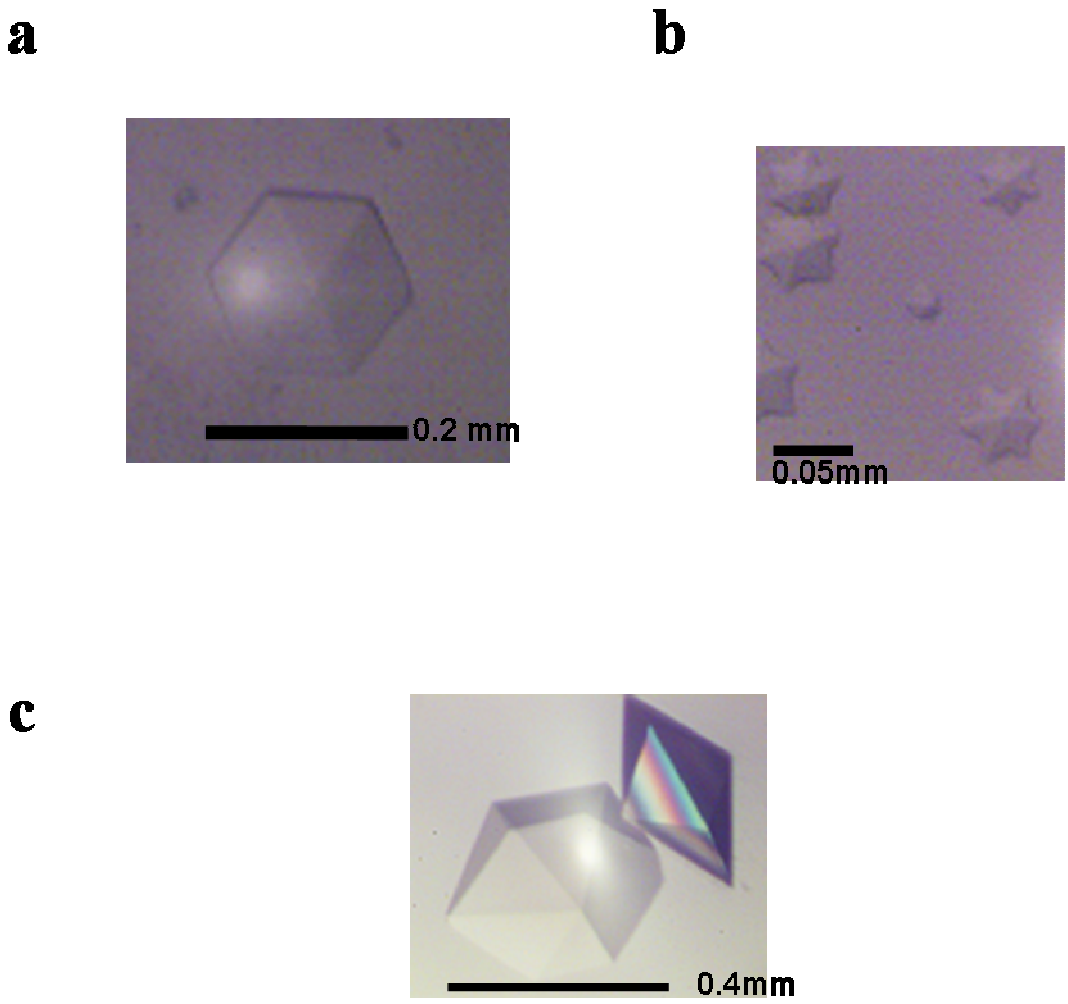


Figure 3.1 Crystals of Apo and Cu(I) bound CsoR

a. Small pellet-like crystal formed by Apo:CsoR is shown. b. Tiny crystals of Cu(I):CsoR from initial screens are shown. c. Large crystals of Cu(I):CsoR obtained after optimization are shown.

(13.3%) of the accessible surface area between the dimers and does not involve any conserved residues at the interface; hence, it may not be of functional importance."

Secondary structure prediction for CsoR was done using different PSIPRED (**3.4.1**, http://bioinf.cs.ucl.ac.uk/psipred/). The region from residue 93 to 119 is predicted to be a coiled region with no α-helical or β-strand elements. Although residues 90, 91 and 92 are predicted to be α-helical, the confidence associated with this prediction is relatively low. This is consistent with our crystallographic results, since regions that form random coils are often flexible and very flexible regions may not contribute to X-ray diffraction. The crystal structure matched the predictions from PSIPRED.

The crystallographic asymmetric unit composed of a protomer of CsoR is shown in **Figure 3.2**. The dimer generated due to crystallographic 2-fold symmetry is shown in **Figure 3.3**.

Each protomer of CsoR contains three α-helices (α1, α2 and α3). α1 spans residues 7 to 32, α2 spans residues 36 to 63 and α3 spans residues 75 to 86. α1 and α2 are connected by a 3 residue loop L1 (residues 33 to 35) and α2 and α3 are connected by a 10 residue loop L2 (residues 64 to 74). The secondary structure elements seen in the crystal structure have been marked on the CsoR sequence (**3.4.1**). The α1-α2 helices of one protomer pack against α1' and α2' of the protomer generated by symmetry. Together, α1-α2 and α1'-α2' form a compact antiparallel four-helix bundle (**Figure 3.3 and Figure 3.4**). Along the top of the four-helix bundle are multiple salt-bridge interactions, for example, Lys17-Glu31', Arg20-Asp24', Asp24-Arg20' and Glu31-Lys17'. The core of the bundle is lined with hydrophobic residues (**Figure 3.3b**). The residues Leu13, Leu16, Val19, Leu23, Val27, Leu30 on the α1 helix and Val39, Val46, Leu50 on the α2 helix from each protomer point towards the core of the four-helix bundle. The core is likely to be stabilized by hydrophobic interactions between the α1-α2 helices of the same protomer as well as between opposite protomers. Stabilization of long α-helices by hydrophobic interactions along the core, is reminiscent of coiled-coil proteins with the leucine zipper motif .

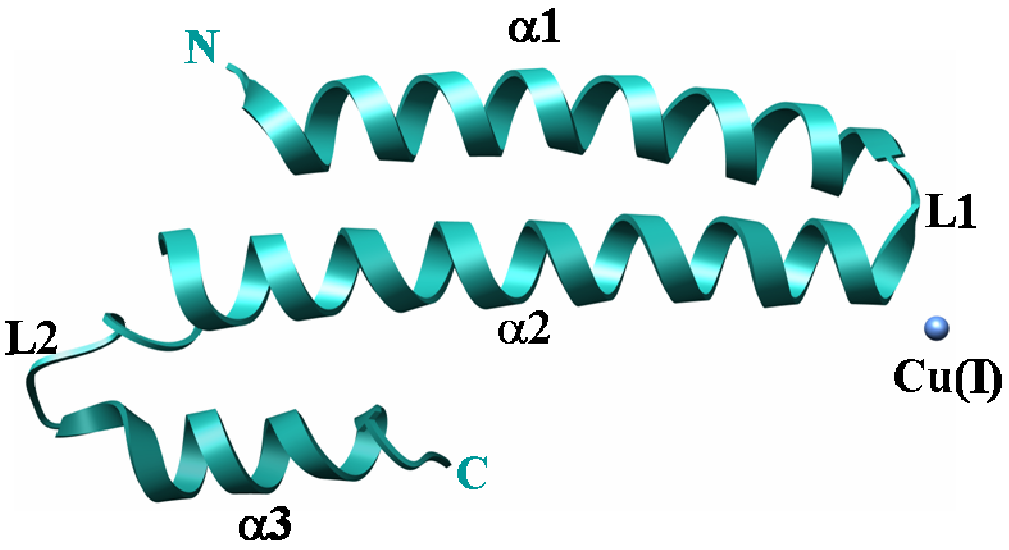


Figure 3.2 The crystallographic asymmetric unit of Cu(I):CsoR

The asymmetric unit of CsoR consists of one protomer of CsoR. Each protomer contains 3 helices- α1, α2 and α3 and 2 loops L1 and L2 connecting the helices. A single Cu(I) ion (blue sphere) is present close to a cysteine residue (Cys36) from the α2 helix..

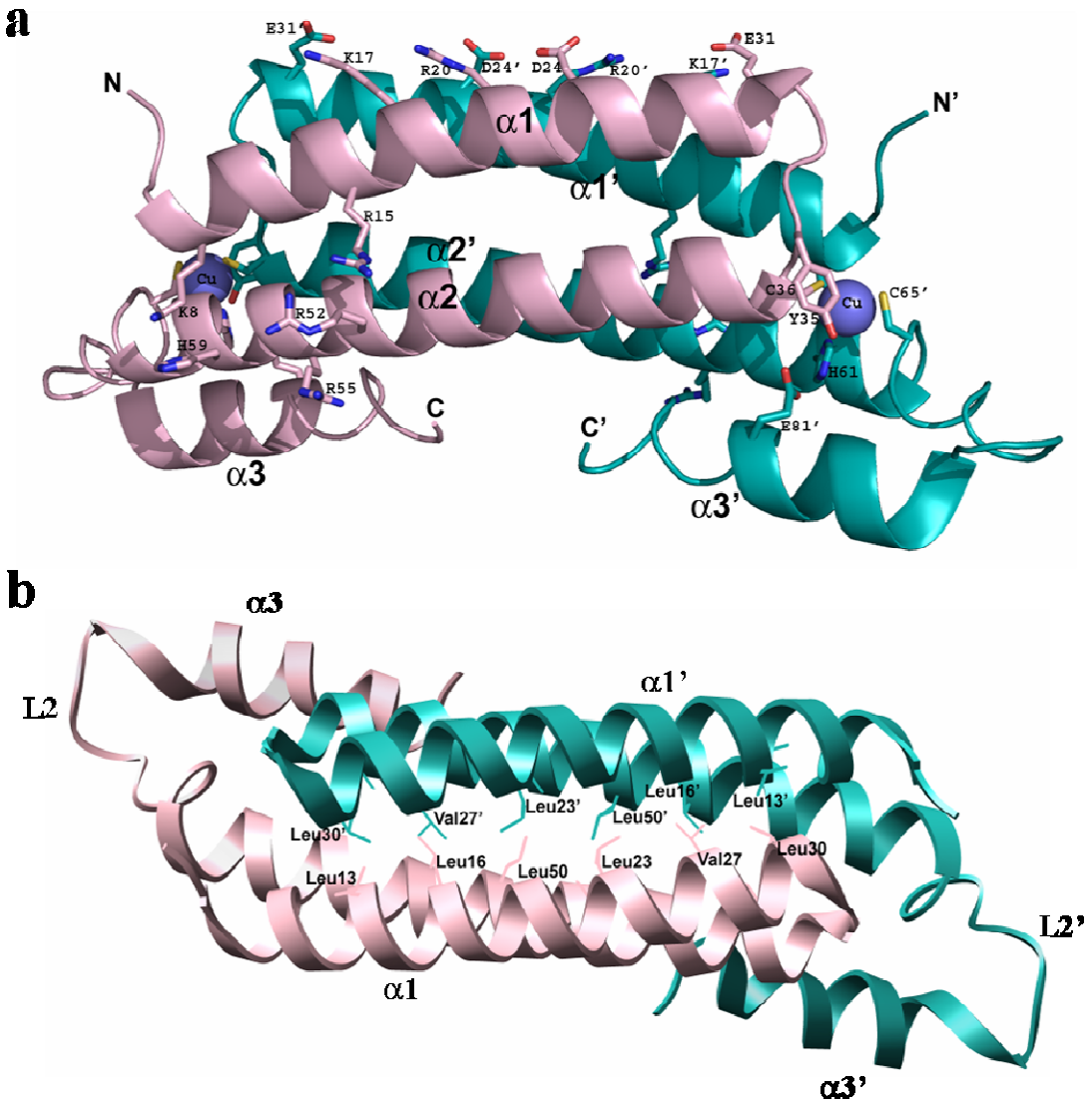


Figure 3.3 Overall crystallographic structure of Cu(I) bound CsoR

a. CsoR is an α-helical dimer of 2 protomers (pink and teal). A single protomer (pink) and is composed of 3 helices- α1, α2 and α3. The α1, α2 of one protomer and α1', α2' of the other protomer together form a 4-helix bundle stabilized by salt-bridge interactions between pairs of oppositely charged residues of α1 and α1'(K17-E31' , R20-D24', D24-R20' and E31-K17') shown as sticks. A cluster of positively charged residues (K8, R15, R52,R55 and H59) is shown as sticks on the front and back face of CsoR. Cu(I) atoms are shown as blue spheres, coordinated by C36, C65' and H61' shown as sticks. Residues Y35 and E81', shown as sticks, are close to the active site. b. Top view of the dimer shows the four helix bundle formed by the α-helices α1 and α2 of each protomer (pink and teal). Hydrophobic residues lining the core of the four-helix bundle are shown as sticks.

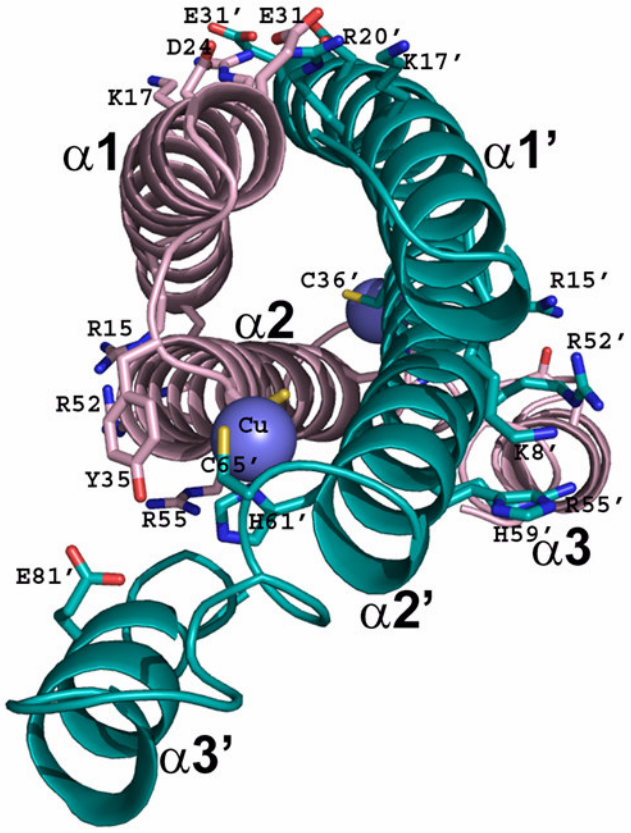


Figure 3.4 View along the long axis of the Cu(I) CsoR dimer

Viewing along the long-axis of the CsoR dimer (pink and teal) shows a closely interacting, compact 4-helix bundle with the two copper-binding sites on either end. The two clusters of positive residues can be seen on the outer face of the bundle. Residues E81 and Y35 of opposite protomers close-in around the metal site.

The α3 and α3' helices are placed on either end of the long axis of the four-helix bundle. α3 packs against the residues at the beginning of α2' helix of the other protomer (residues 37 to 44). The α3-α2' interface is small and most pairs of residues are greater than 3.7Å apart.

A surface representation of the CsoR dimer colored by electrostatic potential (**3.4.1, Figure 3.5**) shows one cluster of positively charged residues. The main contribution to this positive patch comes from residues Arg15 and Lys8 from the α1 helix and residues Arg52, Arg55 and His59 from the α2 helix of the same protomer. Owing to the 2-fold rotational symmetry in the CsoR structure, a corresponding positive patch is seen on the other face of the dimer. This can be seen while looking at the dimer from the bottom, as shown in **Figure 3.5.**

3.2.2 The Cu(I) site in CsoR

As explained below in section 3.3.5, electron density corresponding to positive 4σ contour was observed close to Cys36 at the N-terminus of the α2 helix. Since the X-ray fluorescence scan performed on the crystals had confirmed the presence of copper (see **section 3.3.4**), a Cu(I) ion was refined into this density. After applying symmetry to generate the dimer of CsoR, the Cu(I) ion was surrounded by Cys36 of the α2 helix of one protomer, His61' from the C-terminus of the α2' helix of the other protomer and Cys65' from the L2' loop of the other protomer (**Figure 3.6**). A second Cu(I) site related by 2-fold symmetry is formed by Cys36' of α2', His61 from α2 and Cys65 from the L2 loop. As described by Liu et al. (**3.4.1**), "The Cu(I) ion is bound to an intersubunit coordination site formed by the Sγ atoms of Cys36 and Cys65' (of the opposite subunit) and the Nδ1 atom of His61'." The Cu(I) ion is 2.66Å away from the Sγ atom of Cys36, 2.68Å from the Sγ atom of Cys65' and 2.4Å away from the Nδ1 atom of His61'. The ligands along with the Cu(I) ion exhibit almost perfect trigonal coordination geometry. The bond angles defining the Cu(I) geometry are 123.7º ($S\gamma_{36}$-

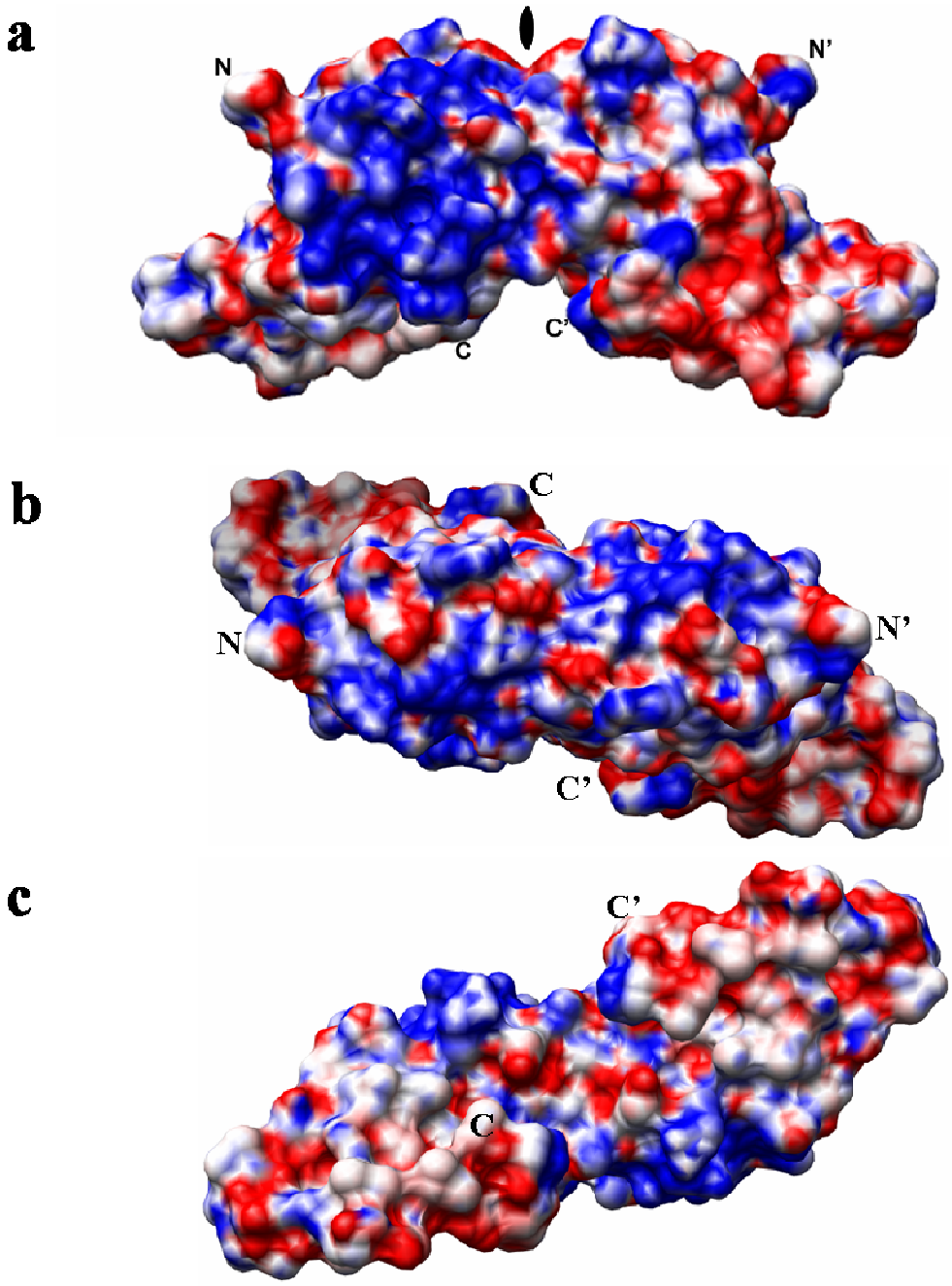


Figure 3.5 Electrostatic surface representation of Cu(I) bound CsoR

a. The electrostatic surface potential rendering of Cu(I):CsoR shows two distinct positively charged surfaces (blue) related by 2-fold rotational symmetry (symmetry axis marked in black) . The negatively charged region is shown in red. b. Top view of the Cu(I):CsoR dimer. c. Bottom view of the Cu(I):CsoR dimer.

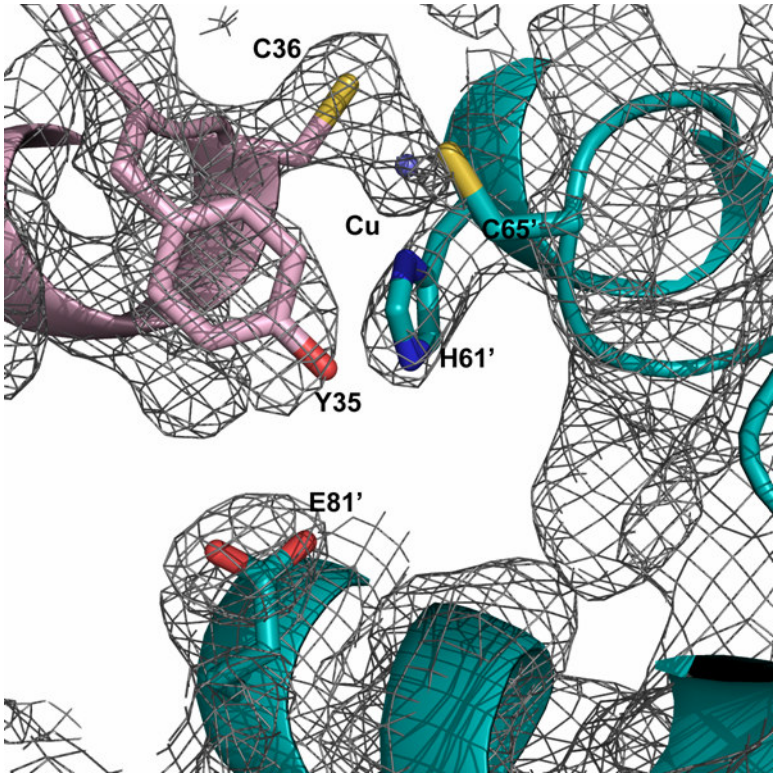


Figure 3.6 The subunit bridging metal site in CsoR

The coordination sphere around Cu(I) in the Cu(I):CsoR structure is shown, with the two protomers in pink and teal. The trigonal metal binding site involves one cysteine (C36) from α2 of one protomer and a cysteine (C65') and histidine (H61') from α2' of the other protomer. In the vicinity of the metal binding site, are residues Y35 and E81' at distances of 3.31Å and 4.28Å respectively, from the nearest ligand residue. Electron density contoured at 1σ , from a composite omit – map is shown as a grey mesh around the residues and Cu(I) atom.

Cu(I)-S$\gamma_{65'}$), 119.8º (N$\delta 1_{61'}$-Cu(I)-S$\gamma_{65'}$) and 116.4º (Sγ_{36}- Cu(I)-N$\delta_{161'}$). The torsion angle between the ligands and the Cu(I) ion is 177.9º which is close enough to a planar geometry with a torsion angle around 180º. Although the bond angles calculated from the Cu(I):CsoR structure match well with the ideal angles for trigonal geometry (around 120º), the metal-ligand distances are larger than typical distances. Distances calculated using Extended X-ray absorption fine structure (EXAFS) analysis of Cu(I):CsoR correspond to 2.21 for the Cu(I)-S bonds and 2.06 for the Cu(I)-N bonds (**3.4.1**). This is consistent with the distances observed in other proteins. Typically Cu(I) prefers tetrahedral, trigonal and sometimes linear geometry. RcnR, a nickel and cobalt responsive metalloregulator from *E. coli* is predicted to have an all α-helical structure (204). Secondary structure predictions suggest a structure similar to CsoR (204). In RcnR, cysteine35 and histidine 60 are likely to represent part of the metal site. Residue 64 varies across the homologs and might affect metal specificity (204). It is possible that structural changes at the metal site involving movement of the ligands helps this three-helical fold to offer trigonal versus square planar geometry. In this way, proteins with similar folds may be specific for different metals.

The role of Cysteine residues as metal coordinating ligands in CsoR was expected, as suggested by S^- to Cu(I) ligand-to-metal charge transfer (LMCT) experiments. The role of His61 however, was first observed in the crystal structure. EXAFS analysis of Cu(I) binding by CsoR confirms the role of His61 is directly coordinating the Cu(I) ion. This is consistent with the ligand preference of Cu(I). Cu(I) is classified as a soft acid based on the ionization potential, ionic radius and ionic charge (135). Soft ions prefer to coordinate to Sulphur (S) or Nitrogen (N) donors as compared to the hard ions like Mg^{2+} and Ca^{2+} that prefer to coordinate to Oxygen (O) donors.

Tyr35, the residue preceeding the metal coordinating Cys36, is in the vicinity of the metal site. Liu et al. describe the changes in the fluorescence properties of this residue upon binding of Cu(I) by CsoR (**3.4.1**). Considering the distance of 3.3Å between Tyr35 and the nearest copper ligand, it is possible that structural rearrangements at the metal site are transmitted to the Tyr35 residue, causing a change in its

fluorescence.

Also, Glu81 a conserved residue from the α3 helix is 4.3Å away from the nearest copper ligand while the Nδ1 atom of His61 directly coordinates the Cu(I) ion. Mutation of Glu81 to Ala decreases the regulation of DNA binding without affecting Cu(I) binding (**3.4.1**). Put together, this suggests that Glu81 could be involved in a conformational switch that may accompany Cu(I) binding, which may decrease the affinity of CsoR for DNA.

3.2.3 Comparison of CsoR structure with other Cu(I) responsive transcriptional regulators

CueR, a member of the MerR family of metalloregulatory proteins, regulates the transcription of genes encoding CopA and CueO which presumably constitute the primary copper export system in *E. coli.* CueR is a homodimeric protein (**Figure 3.7,** (176), (14) that binds 3 equivalents of Cu(I) per monomer. One Cu(I) site is formed by residues Cys112 and Cys120 from the same protomer. These residues lie in a 10 residue loop which turns around to bring the two cysteines together to form metal coordinating site. Cu(I) is bound in a 2-coordinate, linear geometry at this site. Each cysteine residue is 2.14Å away from the Cu(I) ion. This metal site represents a tight binding site relative to two other weak metal binding sites close to Cys129 and Cys130 that are present in the C-terminus of the protein. The C-terminus of the protein was disordered and could not be resolved in the crystal structure. It is speculated that the unusual linear 2-coordinate geometry might contribute to the metal specificity of CueR, allowing it to bind only Cu(I), Ag(I) and Au(I), not Zn(II) or Hg(II). The Ag(I) and Au(I) bound crystals structures of CueR show an almost identical metal site as that bound to Cu(I) except for an increase in the metal sulfur bond distances. This idea is consistent with the features of other metalloproteins in which metal specificity is determined by the coordination and geometry of the metal site (174).

CsoR shows obvious structural differences from CueR. CsoR possesses only 2 cysteine residues, both of which take part in metal coordination. Each Cu(I) site is a

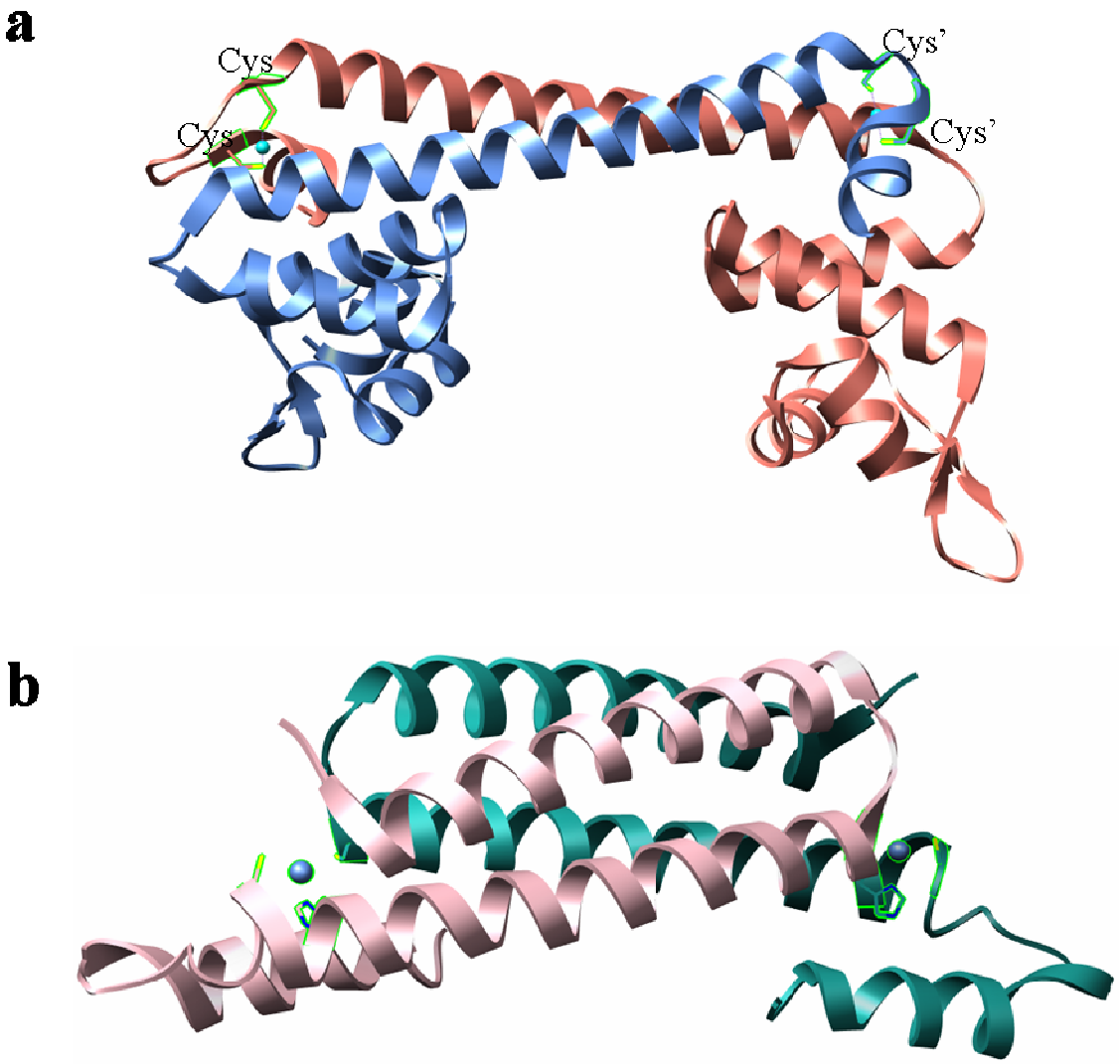


Figure 3.7 Comparison of metal sites in CueR and CsoR

a. Crystal structure of CueR bound to Cu(I), PDB 1Q05. CueR forms a homodimer (salmon and blue). A single Cu(I) (cyan sphere) is coordinated in a two coordinate linear geometry by two cysteine residues from the same protomer. b. In CsoR, Cu(I) is present in a subunit bridging site with ligands from both protomers.

subunit bridging site since ligands from both protomers of the homodimer are required to coordinate the Cu(I) ion. This is in contrast to the Cu(I) site in CueR. Although the metal site is formed at the dimer interface, both cysteines that coordinate Cu(I) belong to the same protomer. Whether in the case of CueR or CsoR, the dimer formation is apparently not driven by metal binding since the apo proteins also exist as homodimers. However, the location of the metal ion at the subunit interface could help in mediating the conformational changes that accompany the switch between metal bound and DNA bound states. In SmtB, a Zn(II) binding transcriptional regulator it has been shown (205) that rearrangement of a complex hydrogen bonding network between the subunits is likely to drive the structural arrangements that take place between the metal bound and DNA bound forms of the protein.

3.2.4 DNA binding fold of CsoR

Metalloregulators have been classified into six structural families. Structural and biochemical studies have shown that metal specificities, coordination geometries, ligand types and number of ligands differ between and even within each family. However, the DNA binding domains are highly similar in structure across at least five families and mostly consist of the winged helix motif or modifications of the helix-turn-helix motif. The winged helix motif is an α/β type of motif with three α-helices (H1, H2 and H3) and three β-strands (S1, S2, S3). Two large loops called wings (W1 and W2) connect S2 and S3 strands and lie on either side of helix H3. In this respect, the structure resembles a winged helix (206). The main distinction between winged helix motifs and helix-turn-helix motifs is the length of the turn connecting helices H2 and H3. This turn is much longer in the winged helix domain. H3 is called the recognition helix and makes contacts with the major groove of the DNA via conserved motifs (Arg-Arg-X-Tyr-Asp). The role of the W1 and W2 wings differs with protein. In CueR, as in all other members of the MerR family, an N-terminal winged helix domain interacts with DNA where as the effector binding site is at the C-terminus (**Figure 3.8a**). Although apart in primary structure of the protein, due to the homodimeric arrangement of the protein, the DNA

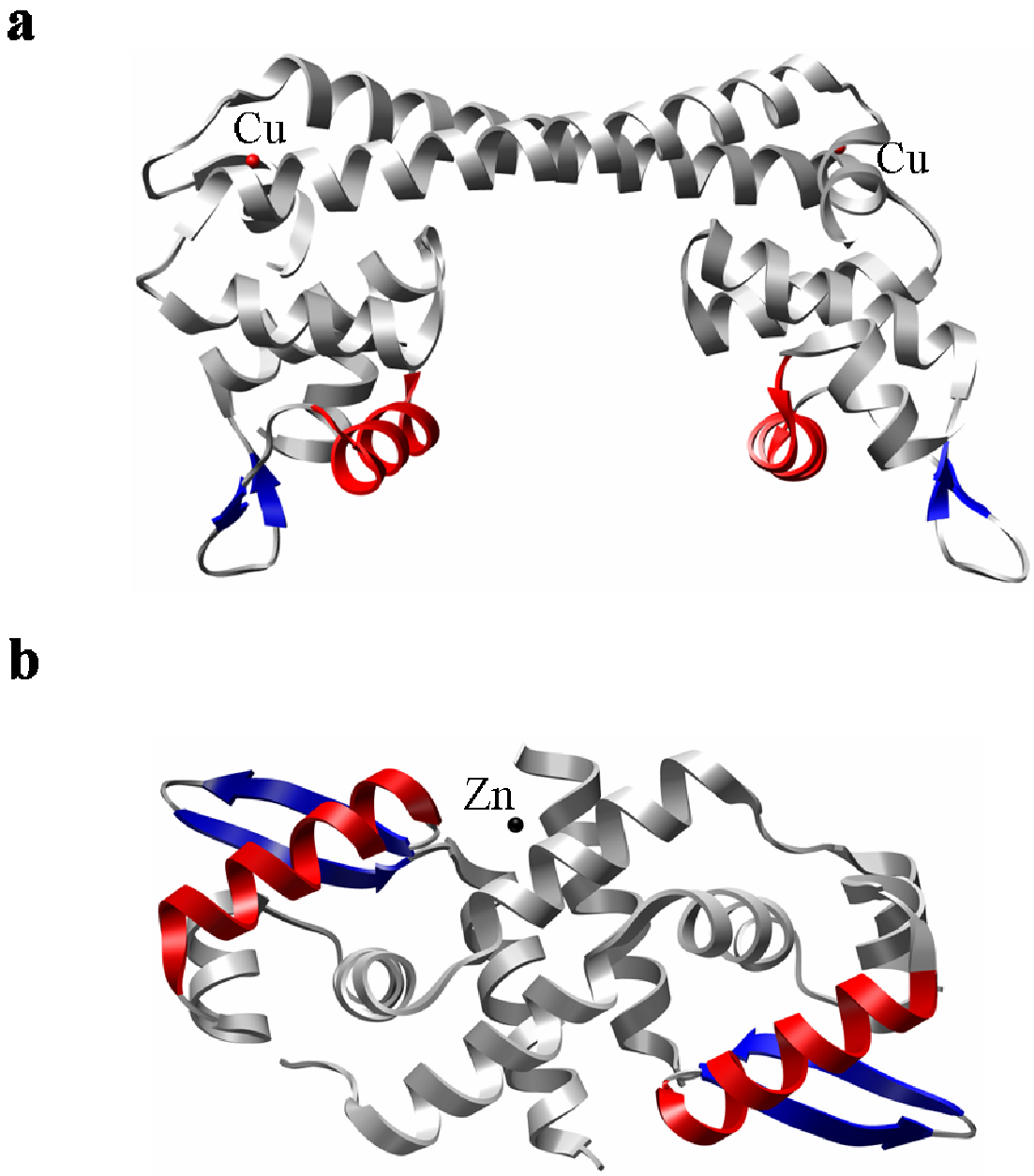


Figure 3.8 DNA binding domains in MerR and ArsR family metalloregulators

a. Crystal structure of Cu(I) bound CueR (1Q05) shows a winged helix motif with the recognition helix shown in red. b. Crystal structure of Zn(II) bound SmtB (1R23) shows a winged helix-turn-helix motif with the recognition helix shown in red.

binding N-terminus and effector binding C-terminus are brought together in close proximity to each other. In SmtB, a Zn(II) binding ArsR/SmtB family member, a winged helix-turn-helix is the putative DNA binding site protein (PDB ID 1SMT), (**Figure 3.8b**). Structures of DNA bound metalloregulators are available for 2 members of the DtxR and 1 member of the NikR family (184), (15), (23). As shown in **Figure 3.9**, while DtxR type regulators use a winged helix motif to recognize DNA, NikR uses a ribbon-helix-helix motif to interact with DNA.

In the CsoR homodimer there are no recognizable DNA binding motifs like the winged helix or helix-turn-helix motifs. We speculate that the positive patch of residues seen on the electrostatic surface of the dimer along with the 2-fold rotation related patch of positive residues represents at least part of the DNA binding site. Consistent with this idea, alanine substitutions of the residues Arg15 and Arg52 that contribute to this positively charged patch, lose the ability to bind DNA as effectively. The positively charged surface potential is likely to aid in charge neutralization of the highly negative phosphate backbone of the DNA. Also, a 2-fold related site could accommodate a pseudo-palindromic sequence as that recognized by CsoR. The C-terminus of CsoR was not resolved in the crystal structure. Although absent in other CsoR homologs, it is possible that this region contributes to DNA binding in *M. tb* CsoR. One possible DNA binding site could involve the top of the CsoR homodimer along with the 2-fold related sites. Another possibility is that the DNA could bend around the bottom of the CsoR dimer to contact the 2-fold related positive patch consisting of Arg15 and Arg52. The distance between the C-α positions of the two Phe88 residues in the homodimer of CsoR measures ~26Å. This is in the range that could accommodate the width of a DNA double helix. These models for DNA binding are currently being tested. Although the DNA binding site of CsoR is not yet well characterized, structural comparisons suggest that CsoR possesses a DNA binding site different from other known metalloregulators. In this way CsoR represents a novel class of metalloregulators.

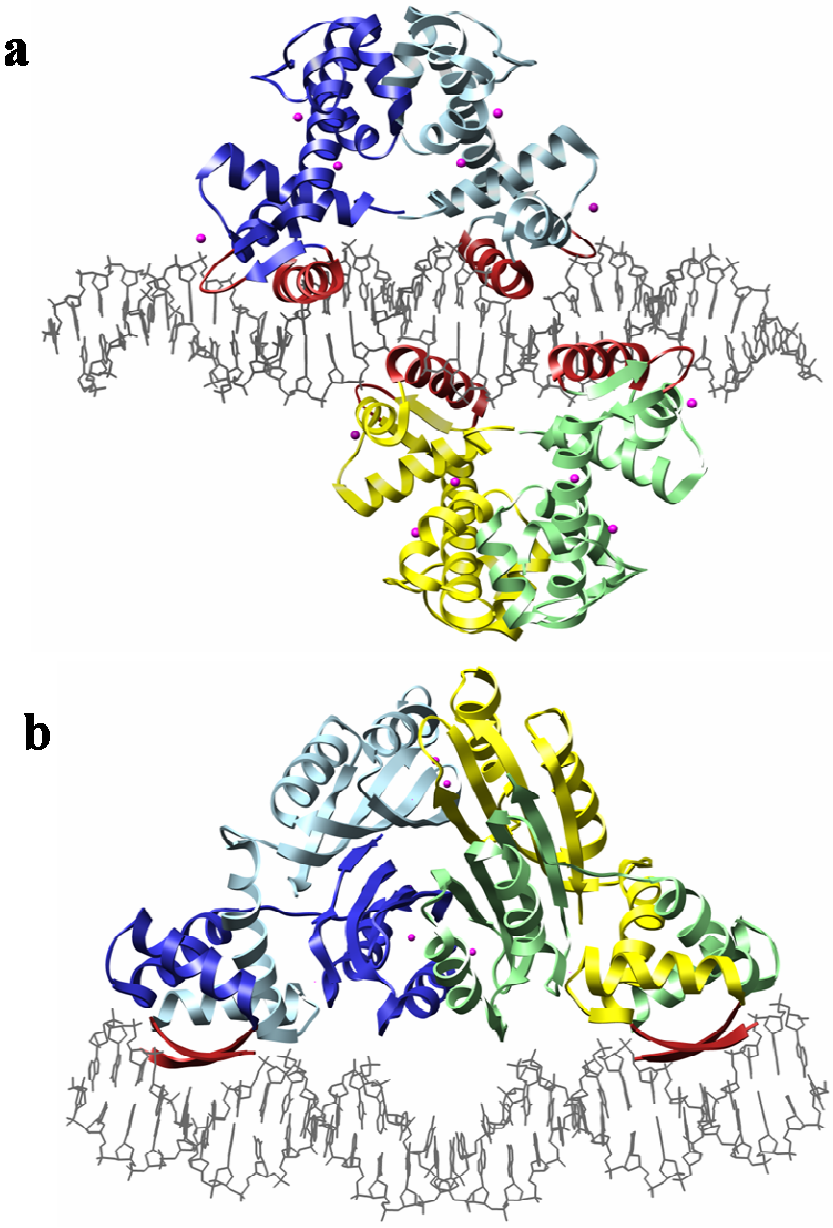


Figure 3.9 DNA binding by DtxR and NikR family metalloregulators

a. Crystal structure of iron sensing regulator IdeR from *M. tb* in complex with DNA (2ISZ) shows a dimer of dimers. Each protomer recognizes the major groove of DNA (grey) using a recognition helix (red) from a winged helix motif. The first wing regions (red) also interacts with the DNA. Iron is shown as magenta spheres. b. Crystal structure of nickel sensing regulator NikR from *E. coli* in complex with DNA (2HZV) shows a homotetramer. A β-strand from each protomer comes together to form 2 β-sheets in the ribbon-helix-helix motif. 2 such motifs (red) per homotetramer interact with major grooves of the DNA (grey). Nickel in magenta.

3.2.5 Crystallization and structure determination of CsoR:DNA complex

As described by Liu et al. (**3.4.1**), CsoR binds a pseudo-palindromic sequence upstream of the *cso* operon. This region resembles a consensus Mtb promoter sequence. To understand the structural basis of the interactions between CsoR and DNA, co-crystallization trials were set up. Various DNA oligos were used including a 20mer, a 24mer, a 28mer and a 33mer encompassing the palindromic promoter region. Details are described in **section 3.3.6**. For initial crystal trials, the DNA oligos were annealed to form a double stand and then mixed in equimolar ratios with Apo:CsoR at a concentration of 420µM. Crystal screen 1 and 2, Wizard screens 1 and 2 and the Natrix screen were used for the trials. Crystals for the CsoR:33mer were obtained within a week in Wizard screen 1 #4, a condition consisting of 2-methyl-2,4-pentanediol, imidazole pH8.0 and magnesium chloride. During optimization to improve the crystal quality, drops were also set up using CsoR alone and the 33mer DNA alone as controls. Within a week, the drops containing DNA alone also showed crystals identical to the original crystals from the crystal screen. Therefore it was concluded that the crystals belonged to the 33mer DNA alone and not the protein:DNA complex.

The affinity of CsoR for the promoter DNA is in the µmolar range (**3.4.1**). It is possible that this relatively low affinity complex is not stablilized enough in the crystallization experiments, making it difficult to obtain crystals of the complex. Also, given the right conditions, the DNA being a relatively stable double-helix structure may nucleate on its own, shifting the equilibrium away from complex formation in the crystallization drops. It is also possible that slow oxidation of the protein occurs in the course of the crystallization experiment, further decreasing the affinity of CsoR for DNA. One approach to solving this problem could be to stabilize the reduced state of CsoR using Cu(I) chelators or better reducing agents. Also, rational design of DNA oligos which are bound with higher affinity by CsoR could potentially increase the chances of obtaining crystals for the complex. Other methods like NMR and fluorescence based techniques might prove more successful in obtaining structural

information about CsoR:DNA binding.

3.3 EXPERIMENTAL APPROACH AND DESIGN

3.3.1 Purification of Apo, Cu(I) bound and Se-met CsoR

Purification of Apo and Cu(I) CsoR has been described by Liu et al. (**3.4.1**). "To obtain phase information for structure determination, seleno-methionine (Se-met) substituted CsoR was produced in *E. coli* B834 (DE3) cells (Novagen). The cells were transformed with the pET3a-CsoR construct, grown in LB media to mid-log phase, transferred to M9 minimal media supplemented with L-amino acids (Sigma, 50mg/L), L-seleno-methionine (Acros, 50mg/L), glucose (0.4% w/v), calcium chloride 0.1 mM), magnesium sulfate (4 mM), thiamine (5µg/mL) and ampicillin (50 µg/mL), induced with 0.5 mM IPTG and allowed to grow for 17 h at 25 ˚C. Following harvesting of culture, the same purification protocol was followed as outlined for the native CsoR." The final purity of the sample used for crystallization was ~70% but despite the low purity, this sample yielded crystals.

3.3.2 Crystallization of CsoR

Apo CsoR at a concentration of 8mg/ml was used for initial robotic screening using Crystal screens 1 and 2 and Wizard screens 1 and 2. Intelli plates were set up with 100µl reservoir solution in a sitting drop method. Apo CsoR crystallized within a week in Crystal screen 1 #39 (0.1M Hepes pH 7.5, 2%v/v PEG 400, 2M ammonium sulfate) and subsequently in many different conditions. Extinction of polarized light can be used as a test to determine if material is crystalline or amorphous. Crystalline material is anisotropic, with different axes having different refractive indices. Polarized light splits up and travels at different speeds through the crystal because of the different refractive indices along different axes in the crystal. If another polarizer is placed near the eye-

piece of the microscope, at certain angles between the crystal and the source polarizer, the net light passing through the crystal has the same polarization as the incident light and gets cancelled by the second polarizer. This is called *extinction of polarized* light (www.hamptonresearch.com). As the second polarizer or crystal is turned to another angle, light starts to pass through the crystal again. Amorphous material is isotropic and does not produce this effect. CsoR crystals caused extinction of polarized light.

To test if the crystals obtained were protein or small molecule crystals, several methods were employed. First, the crystals were stained with the Izit Crystal Dye (www.hamptonresearch.com) which stains protein crystals blue, by entering the large solvent channels. Small molecule crystals because of their close packing do not get stained. The Apo:CsoR crystals took up the stain only after almost 1 day. Crystals of small molecules most often are hard, and crack with a characteristic sound as compared to protein crystals which are soft and can be cut into. Apo:CsoR crystals, when subjected to this "crack" test provided confusing results. Lastly, the crystals were exposed to X-rays to obtain a diffraction pattern. Characteristic protein spots were seen for Apo:CsoR crystals, confirming that the crystals were protein crystals. The difference between diffraction images obtained from protein versus small molecules is shown in **Figure 3.10**. The inverse relation between the distances that separate diffraction spots and the actual distances within the crystal implies that molecules that are packed closely produce spots that are far from each other and vice versa. Small molecules pack closely and hence produce very bright spots at high resolution, with large distances between the spots. Protein crystals produce closely spaced spots since molecules are placed far away in the crystal.

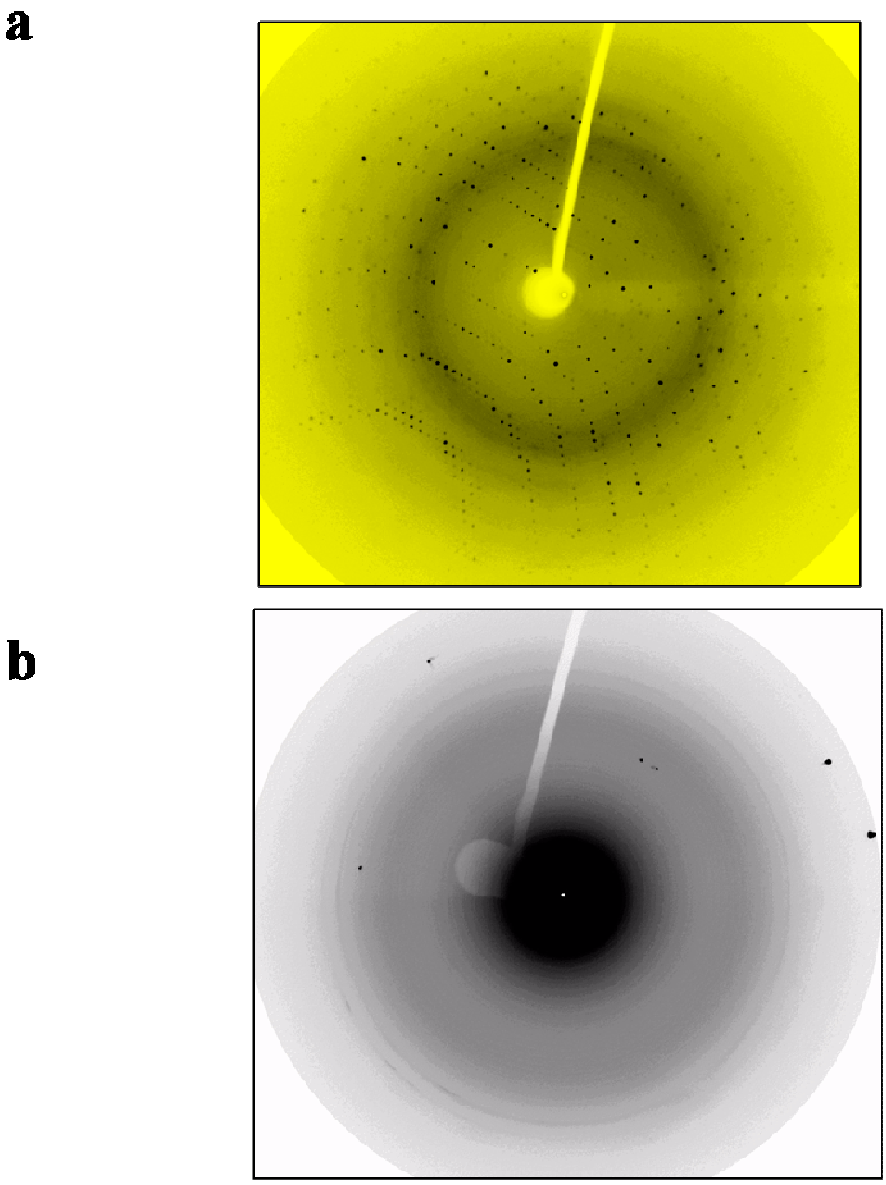


Figure 3.10 Comparing diffraction patterns from protein versus salt

a. Diffraction from a single CsoR crystal shows many spots in the low resolution (centre of frame) and fainter spots in high resolution (periphery of the circle). b. Diffraction pattern from a crystal of magnesium phosphate shows few spots placed far from each other. Few bright spots at very high resolution can be seen.

The Apo CsoR crystals were flat and almost plate-like. To enable the crystals to grow along the third dimension, optimization was done by the hanging drop method. This also facilitated handling of the crystals. To further improve the quality of the crystals, the ratio of protein to crystallization solution was varied. A marked difference was seen as a result of varying drop ratios, with a ratio of 2:4 (crystallization solution:protein) giving much larger crystals (**Figure 3.1c)** with sharper edges as compared to 2:2 or 4:2 ratios. At ratios of 4:2, the drop did not look homogeneous and appeared to show some phase separation. Despite the outward appearance, crystals of Apo CsoR did not diffract well.

Hexagonal crystals for Se-met substituted CsoR were grown by the hanging drop method in a crystallization solution consisting of 0.1 M Hepes sodium pH 7.5, 2% PEG 400 and 1.7 M ammonium sulfate. Diffraction quality crystals were obtained by mixing 2 µL crystallization solution with 4 µL protein at a concentration of 800 µM, and equilibrating against 500 µL of crystallization solution. A marked difference was seen in the size and quality of crystals when the protein was prepared in 25mM MES buffer pH5.9 and 300mM NaCl as compared to 25mM Hepes pH 7.0 and 200mM NaCl.

Crystals for Cu_1 wild-type CsoR were obtained by the hanging drop method in a crystallization solution consisting of 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate pH 5.6, and 30% w/v polyethylene glycol 4000. Diffraction quality crystals were obtained by mixing 2 µL of crystallization solution with 4 µL of protein at a concentration of 400 µM and equilibrating against 500 µL crystallization solution.

SDS-PAGE analysis of the different crystal forms of CsoR was performed. Crystals for each form of CsoR (Apo, se-met Apo and Cu(I) bound) were picked and washed at least twice in mother liquor. This was done to ensure that no contamination occurred from the rest of the material in the drop. **Figure 3.11** shows the migration of the crystals on a Tris-tricine gel (Biorad) after being dissolved in SDS loading dye. The Apo:CsoR crystals and the Cu(I):CsoR crystals ran similar to the full length ~13kDa

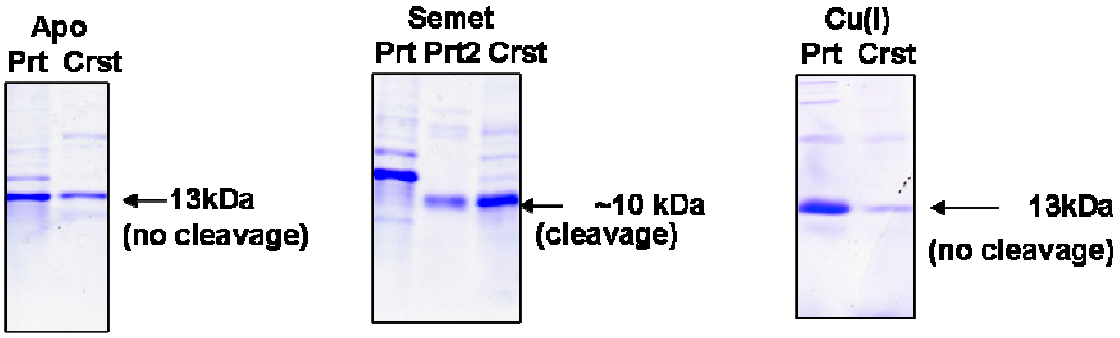


Figure 3.11 SDS PAGE analysis of CsoR crystals

Crystals of ApoCsoR, semet-CsoR and Cu(I) bound CsoR run a 16.5% Tris-Tricine/Peptide gel and stained with Coomassie BB are shown. Apo and Cu(I):CsoR crystals were intact whereas semet-CsoR was missing the C-terminus unstructured tail.

protein used as starting material for the crystallization. The Se-met derivatized Apo:CsoR crystals appeared to be ~10kDa in size. This phenomenon was not observed only in the crystallization experiment since protein left in solution also showed cleavage as shown in **Figure 3.11.**

3.3.3 Preparing crystals for data collection

As discussed in **Chapter II, section 2.3.7**, preparing crystals for diffraction experiments in liquid nitrogen requires the crystal to be soaked in a cryoprotectant that prevents freezing. Crystals of CsoR were soaked in many different cryoprotectants and exposed to X-rays. Some of the cryoprotectants tried were- 15 to 30%Peg 400 in mother liquor (ML), 15 to 30% Glycerol in ML, 20 to 30% ethylene glycol in ML and paratone. The best diffraction was obtained in paratone.

3.3.4 X-ray data collection and data processing

The best crystal of Apo CsoR diffracted to 3Å at the 23ID beamline at Argonne National

labs, IL. The best Semet-Apo CsoR crystal diffracted to 2.13 Å at beamline 8.3.1 at the Advanced Light Source, Lawrence Berkeley National Laboratory. Data statistics are compiled in **3.4.1**. Consistent with the hexagonal morphology of the crystals, the crystals showed 6-fold symmetry and belonged to the P 6_4 2 2 space group with cell dimensions of a = 91.31 Å, b = 91.31 Å, and c = 47.87 Å. The best Cu(I):CsoR crystal diffracted to 2.55 Å at beamline 19-ID at the Advanced Photon Source, Argonne National Labs, Illinois and was isomorphous to the Se-met CsoR crystals with cell dimensions of a = 91.06 Å , b = 91.06 Å, and c = 46.78 Å. Completeness of the diffraction data, defined as ratio of the number of unique reflections measured to the total number of unique reflections, was equal to 99.4% for the overall data and 99.2% at a resolution of 2.55Å. X-ray fluorescence scan of the Cu(I):CsoR crystal showed an emission peak at 8.1 keV (**Figure 3.12**) confirming the presence of copper in the crystal. As expected, a similar scan performed on a Se-met Apo CsoR crystal showed no peak at 8.1 keV.

After data processing, symmetry was estimated using Xprep and HKL2000 (122). The crystals showed P6 symmetry. P6 refers to a hexagonal unit cell characterized by unit cell dimensions of a = b and $\alpha = 90^o$, $\beta = 90^o$ and $\gamma = 120^o$ (117), (118). In crystals with P6 symmetry, there is 6-fold rotational symmetry parallel to the axis along c. A 6- fold gives rise to 6-fold symmetry both on the zero level and on subsequent levels. This can be used to distinguish between 3-fold and 6 –fold symmetries since in the case of 3-fold symmetry, the subsequent levels do not show 6-fold symmetry (118). Higher symmetries in P6 involve a 2-fold rotational symmetry along an axis perpendicular to the 6-fold axis. This results in another 2-fold rotational symmetry being generated along an axis perpendicular to the already existing 6-fold and 2-fold axes. As described in **Chapter II, section 2.3.8**, combinations of rotational and translational symmetries give rise to screw symmetry. The 6- fold rotational axis can have 3 different screws- 6_1 and 6_5, 6_2 and 6_4 and 6_3 (118). Screw symmetries give rise to absences in diffraction and each screw has a characteristic set of *systemtic absences*. By looking for systematic absences, the screw symmetry can be determined. For CsoR

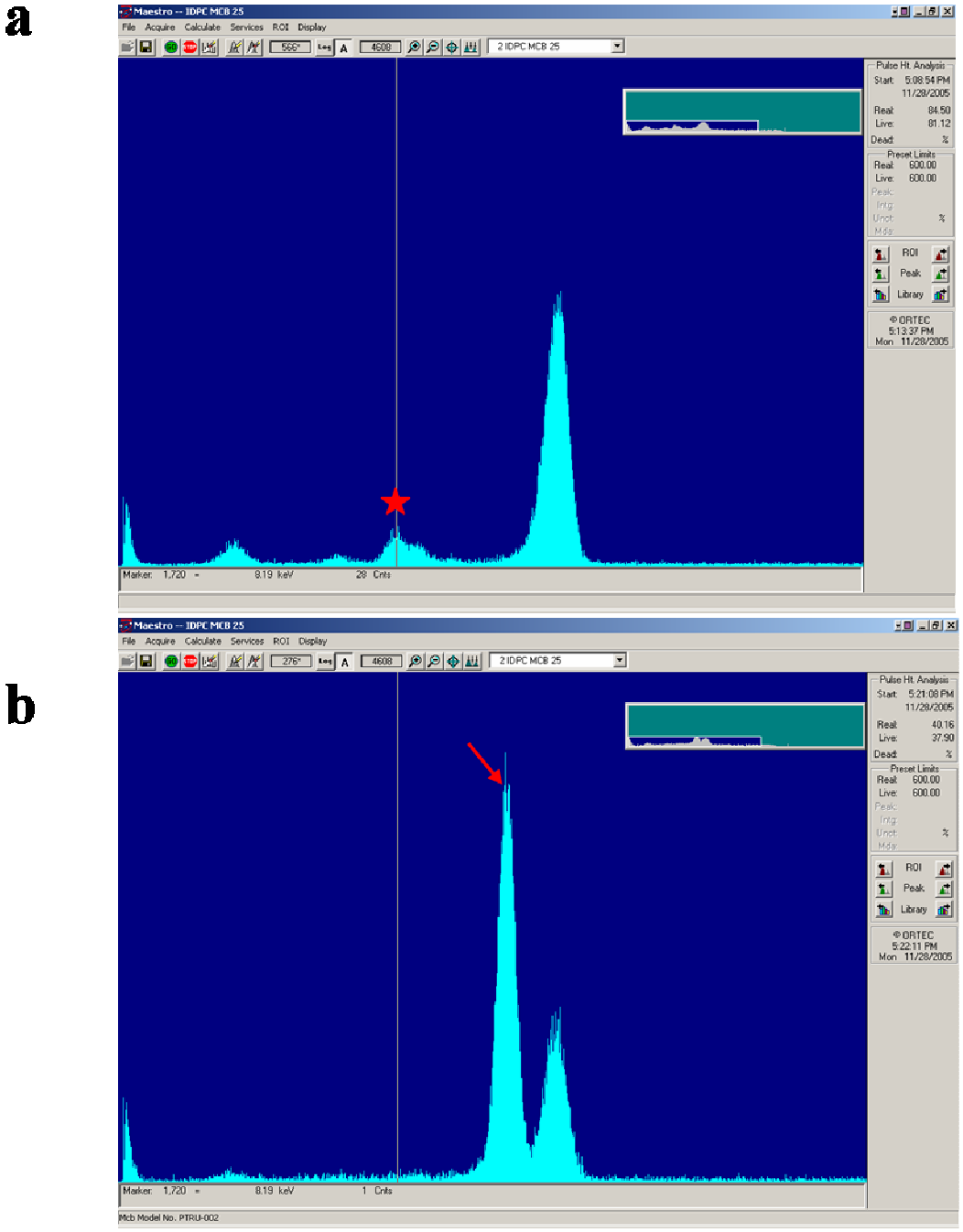


Figure 3.12 X-ray fluorescence scans of CsoR crystals

a. Scan of a Cu(I):CsoR crystal shows a peak (asterisk) at 8.1keV characteristic of copper. b. Scan of a Se-met crystal shows a prominent peak from selenium (arrow) and no peak at 8.1keV from copper.

crystals, both Xprep and HKL2000 gave conflicting results and were unable to distinguish between the higher symmetries of P 6_422, P 6_122, P 6_222, P6_522 and P6_322. To resolve this problem, the precession frames from the diffraction were manually checked along the axes where systematic absences were expected (**Figure 3.13**). The precession frames were viewed using the PRECESS_X program from the PHASES package (207). The expected reflection condition for 6_122 and 6_522 is $l = 6n$ (119). However, in the CsoR diffraction data, the *009 (hkl)* spot is present along the *00l* axis of the *0kl* frame. Also, for 6_322 the reflection condition is l (119), but none of the even spots like the *002, 004* etc were seen. Hence, the space group assigned was 6_422 or 6_222. Both these space groups were used and only in later steps they were distinguishable.

To estimate the number of CsoR molecules in the asymmetric unit, the "Mathews Coefficient" program in the ccp4 suite was used. A single molecule of CsoR was calculated to compose the Asymmetric Unit. The Mathews coefficient, which represents the volume of the crystal per unit mass of the protein, was calculated as 2.2 with a solvent content of 42.4%. This is typical for most proteins, the range being 1.9 to 4.2 (117).

3.3.5 Structure determination of Cu(I)-CsoR

One of the important components of scattered X-rays, called the *phase* is lost in the diffraction experiment (117). This is a required component for calculating the electron density from the diffraction data. Many methods are available to solve the *phase problem*, two of which are described in detail in **Chapter II, section 2.3.9.** Since no homologous structures are known for CsoR, molecular replacement could not be used to calculate phases. Soaking of crystals to obtain phases using *multiple isomorphous replacement* (MIR) is a cumbersome process. Very often, the soaking process disturbs the crystals so that they are no longer isomorphous to the native crystals, and this poses to problem in obtaining phases. Another problem is that a trial and error method is used to find the heavy atoms that specifically bind to sites on the protein. Only after data

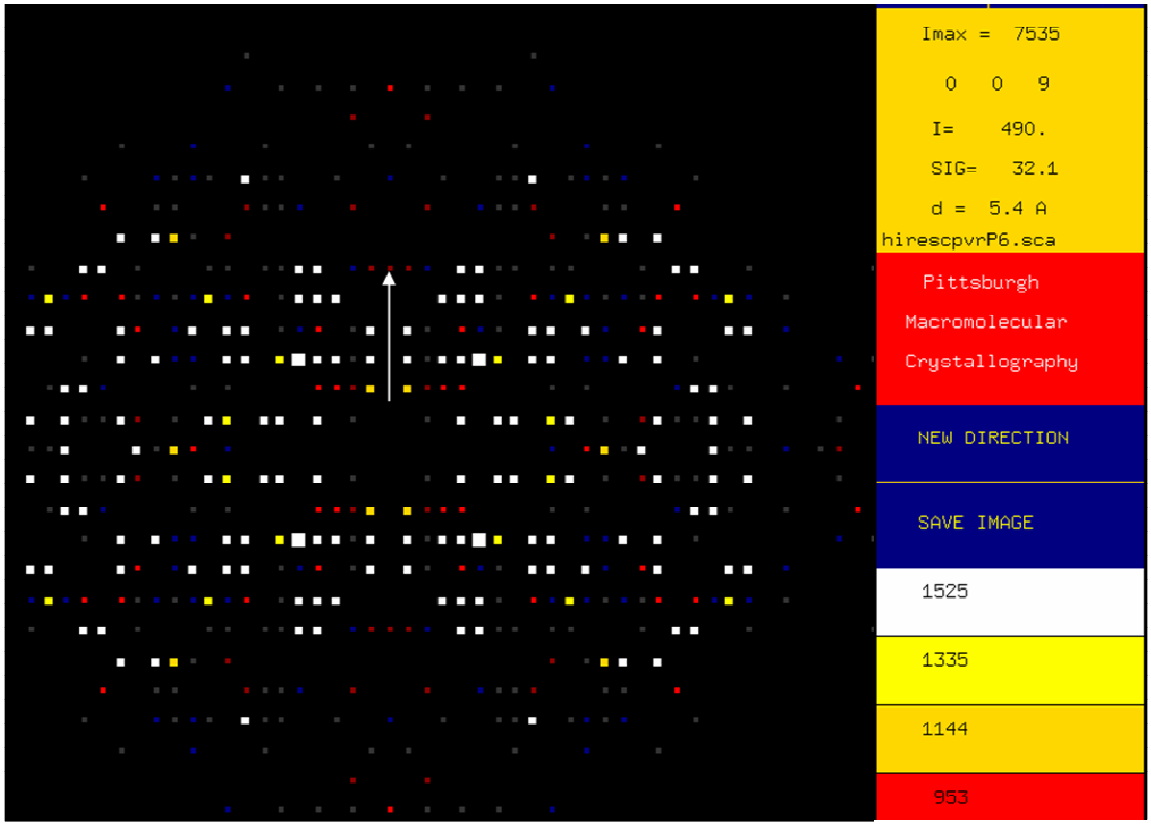


Figure 3.13 Precession frame from CsoR diffraction

Diffraction spots from the *0kl* frame from CsoR diffraction are shown. Along the *00l* axis (white arrow), presence of the *009* spot (white arrow-head) and absence of any even spots (like *002, 004* etc.) confirms the symmetry to be $P6_422/P6_222$.

collection and processing, can the signal from the heavy atom be detected. This makes the MIR method relatively long and tedious.

For obtaining phases for CsoR crystals, we considered two options for *multiwavelength anomalous dispersion* (MAD) experiments. First, the protein could be derivatized with Selenomethionine to use the anomalous scattering by selenium to obtain phases. Second, the Cu(I) ion bound to CsoR in the Cu(I):CsoR crystals could be used to obtain phases since Cu(I) is a relatively heavy atom and is known to display anomalous scattering close to the Cu absorption edge. The added advantage of using Cu to obtain phases was that the diffraction data could be collected at the home-source for X-rays since the anode used to generate X-rays is composed of copper and the resultant radiation is in the range where anomalous scattering from Cu could be seen. Since the se-met crystals were larger and diffracted to a higher resolution, the phasing was done using the data collected at the peak, inflection and remote wavelengths in the selenium absorption spectrum. Details of the MAD data collection and structure solution are described in **Chapter II, section 2.3.9**.

Diffraction data sets were merged, scaled and integrated using HKL2000. From the three wavelength MAD data from Se-met CsoR crystals, 3 out of 5 selenium sites were located using phenix.hyss (121). Refinement of selenium sites and initial phase calculation followed by density modification was done using AutoSHARP (123). With these density modified phases, the electron density map was calculated. Using this electron density as input, the initial model was built by TEXTAL (120,208), an automated model building program. The initial model built by TEXTAL is shown in **Figure 3.14a.** This model was refined against the Cu(I):CsoR data first by rigid body refinement using REFMAC on the CCP4 suite(209) and then by simulated annealing refinement using phenix.refine (121). At every stage of refinement, difference density maps were calculated. Difference density maps are calculated by comparing the electron density calculated from a model to the density map used to build it. For example, the $|F_o|$-$|F_c|$ map is a difference map in which there is positive density at sites where the model is not represented enough and negative density where the model has been placed

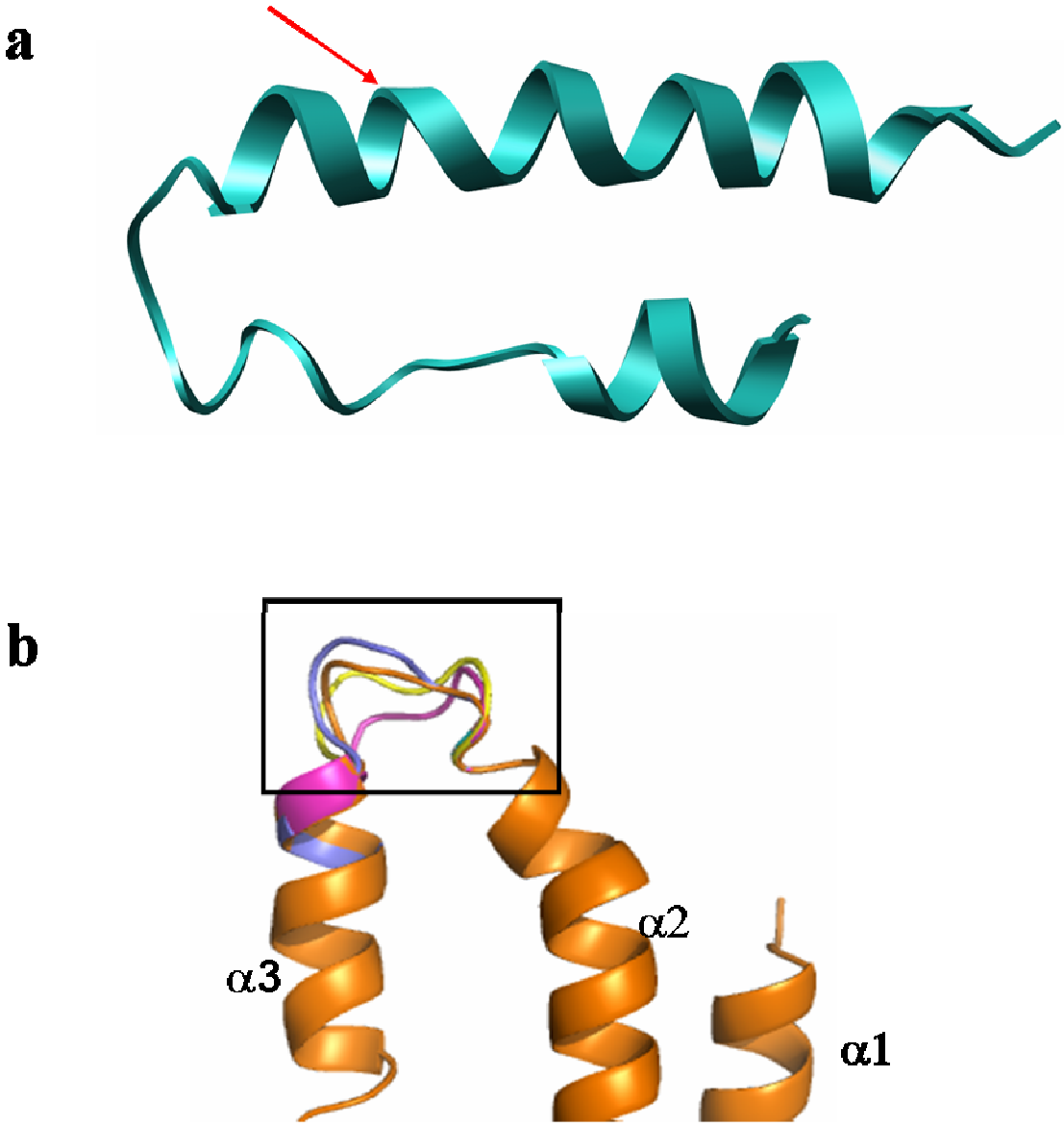


Figure 3.14 Model building of CsoR

a. Initial model of CsoR built using TEXTAL, is shown (teal). Clearly recognizable features like α helices (red arrow) in the initial model are shown. b. Candidate loops (boxed region) built by SwissPDBViewer between helices α2 and α3 are shown in different colors.

wrongly. Using the positive and negative peaks from the $|F_o|-|F_c|$ map, corrections were made to the model. One important contribution from the difference maps was the appearance of positive density for the α3 helix which was not clear in the initial density maps and hence was not built by TEXTAL. Another map with double the contribution from the observed data, called the $2|F_o|-|F_c|$ map was used. Since the α3 helix had not been built so far, by using the map with more contribution from the observed data the density in this region was enhanced. This made it easier to build the α3 helix. Since the initial model built by TEXTAL was using the se-met Apo:CsoR data and the model being built was the Cu(I):CsoR structure, it was possible that the phases being used were biased towards the se-met Apo:CsoR structure. To solve this problem, bias minimized maps were calculated using the TB Bias Removal Server (125). Also, for a more complete removal of model-bias, a simulated annealing *composite-omit map* was calculated using CNS (210) and used as a guide for model improvement. Details of simulated annealing have been discussed in **Chapter II, section 2.3.9.** An *omit map* is calculated after the atoms in a defined small volume are removed and the remaining structure is used to calculate phase angles which are then used to calculate the electron density in the omitted volume (117). A composite omit map uses the principles of an omit map, except that different parts of the model are omitted in successive runs, to cover the entire model. Manual model building using XtalView (124) and refinement using phenix.refine were iteratively done. One of the challenges with building the Cu(I):CsoR model was the loop region between α2 and α3. Density in this region was not very clear. To aid in building this loop, we used the *build-loop* feature in Swiss Pdb Viewer (211). After specifying the residues that define the ends of the loops, the sequence of residues in the loop is specified to the program. A table of possible loops is output, along with the associated energies and a clash score which reports on possible steric clashes. Manual inspection of the suggested loops (**Figure 3.14b)** was done and the best fit was used as a starting point to build the final loop. Positive electron density corresponding to the 4σ contour level was observed in the F_o-F_c map. This density was close to a Cysteine residue (Cys36). Since the X-ray fluorescence scans had confirmed

the presence of copper in the crystal, a single copper atom was refined into this density. Positive densities less than 3σ were identified and water molecules were manually added to these sites to obtain the final model. As described in **Chapter II, section 2.3.9,** *R-factors* were calculated at every step of refinement. The final Cu(I):CsoR model refined with crystallographic R_{work} and R_{free} of 22.63% and 27.62% respectively. Refinement statistics are shown in **3.4.1**. To analyze the geometry of the protein backbone, a Ramachandran plot was made (**Figure 3.15**) with the backbone ψ angle plotted against the φ angle. 97.5% of the residues lie in the allowed region and 2.5% of the residues lie in the generously allowed region.

Accessible surface area was calculated using UCSF Chimera(128,129), Electrostatic surface potential was calculated using APBS (127) and UCSF Chimera; all other figures were prepared using PyMOL (www.pymol.org).

To analyze the crystal contacts made by CsoR in the crystal, and prompted by the biochemical data suggesting a dimeric form of CsoR (**3.4.1**), the symmetry mates were generated. One of the symmetry mates shared the maximum surface area with the original molecule. Also, upon generating symmetry, the Cu(I) ion was now surrounded by Cys36 of the original molecule and Cys65 and His61 of the symmetry related molecule. Since these are possible ligands that coordinate Cu(I), this symmetry mate together with the original molecule are likely to form the biological dimer. Since the Cu(I) ion lies on a 2-fold rotation axis, generating symmetry resulted in the site being allotted 2 ions of Cu(I). To resolve this problem, the site was assigned an occupancy equal to 0.5 during refinement. Generating crystallographic symmetry also gave rise to higher order structures, as shown in **Figure 3.16**.

3.3.6 Crystallization of CsoR:DNA complex

DNA oligos their complementary sequences were purchased from Integrated DNA Technologies with standard desalting purification and at a 10μmole scale. Five oligos of different sizes, representing the promoter region were ordered. These were- a 20 mer (5'CCCACCCCC AGTGGGGTGGG), 24 mer

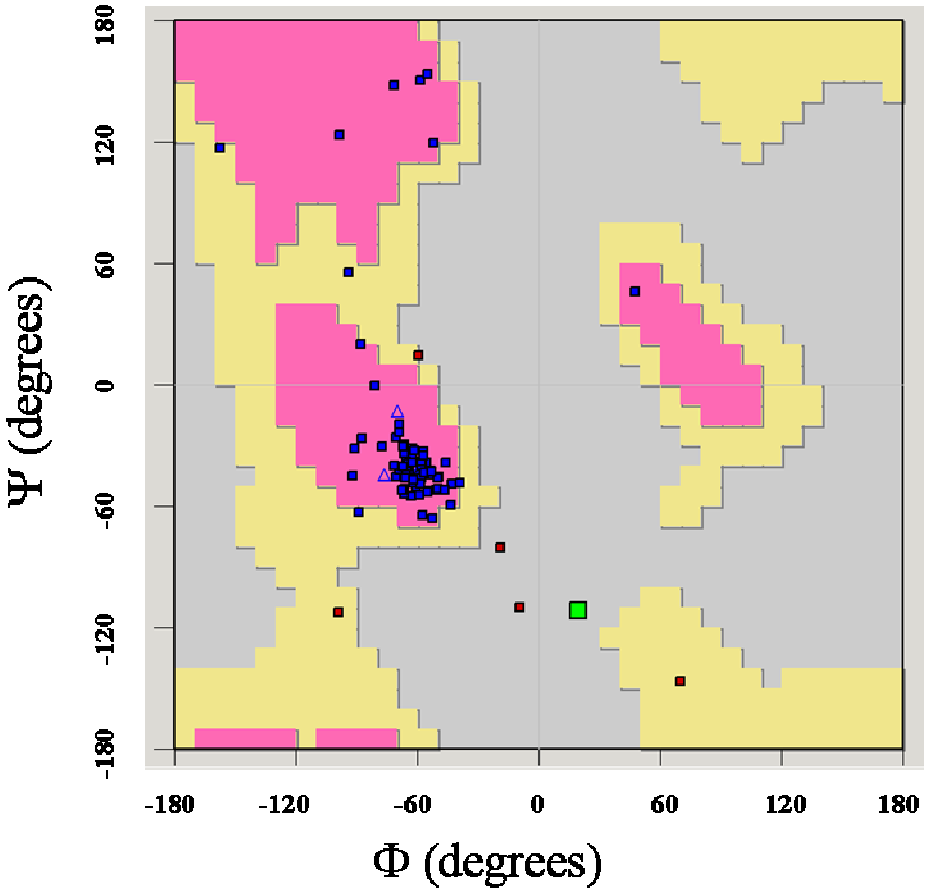


Figure 3.15 Ramachandran Plot for the Cu(I):CsoR crystal structure

Backbone torsion angles of the CsoR structure are shown. Allowed regions are marked in pink and additionally allowed regions are marked in yellow.

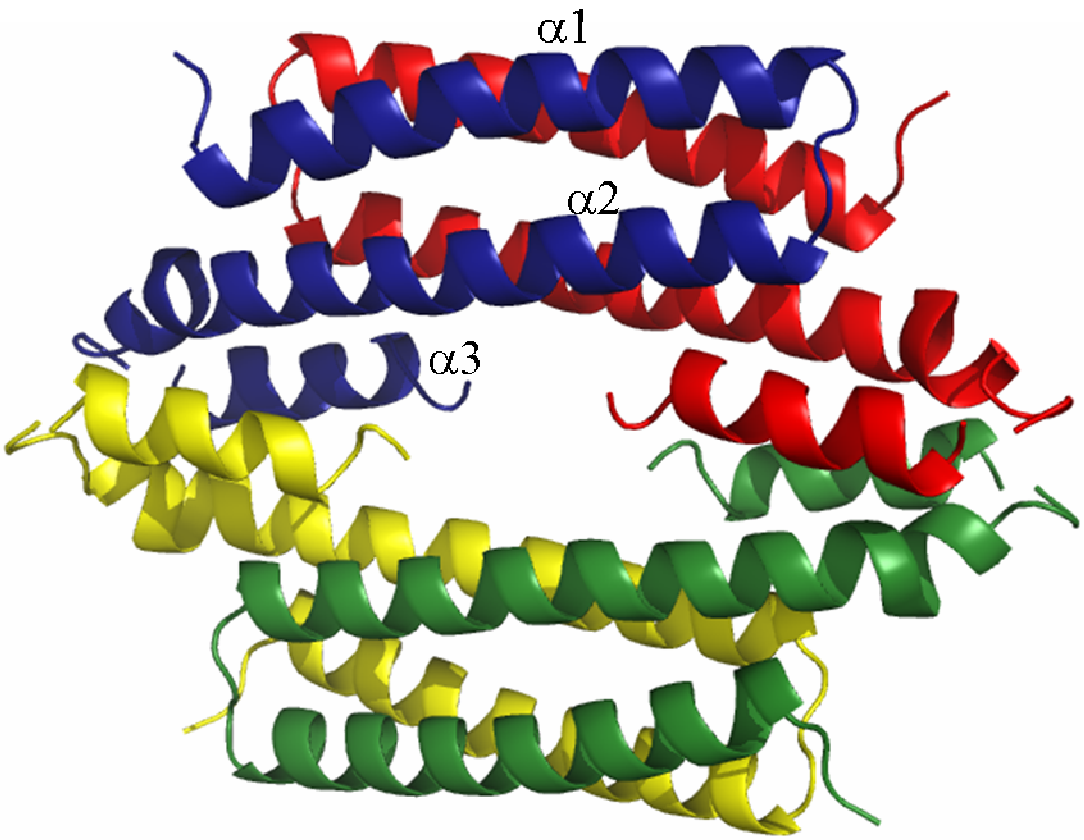


Figure 3.16 Higher order oligomers of CsoR

Crystallographic symmetry mates (red, blue, yellow, green) give rise to a dimer of dimers. The three helices- α1, α2 and α3 of one protomer are marked.

5'AGCCCACCCCCAGTGGGGTGGGAT), 28 mer (5'GT AGCCCACCCCCAGTGG GGTGGGATAC) and a 33 mer (5'GGTAGCCCACCCCCA GTGGGGTGGGATACCATG). The DNA oligos were resuspended in deionized water and annealed at room temperature for 30 mins after heating for 5 minutes at 90ºC. Double stranded DNA fragments were mixed with CsoR at a concentration of 420µM. To remove any aggregates that might have formed after mixing, the sample was centrifuged at 13000rpm on a table-top centrifuge for 10 minutes and used for crystal trials. Crystal screens were set up using a robot, on intelli plates with crystal screen 1 and 2, wizard screen 1 and 2 (Emerald BioSciences) and the Natrix screen (Hampton research). Preliminary crystals grew in 35% 2-methyl-2, 4-pentaneddiol, 0.1M imidazole pH8.0 and 0.2M magnesium chloride. Optimization was performed on hanging drop plates, where drop ratios were varied as done for the CsoR crystals (described in **section 3.3.2**).

3.4 ADDITIONAL MATERIAL

3.4.1 Suggested reading

"CsoR is a novel *Mycobacterium tuberculosis* copper-sensing transcriptional regulator". By Liu et al., *Nature Chemical Biology,* December 2006, is available at DOI:10.1038/nchembio844.

CHAPTER IV
CONCLUSIONS AND FUTURE DIRECTIONS

4.1 STRUCTURAL STUDY OF THE Ro PROTEIN FROM *D. radiodurans*

In this work we have determined a high resolution structure of Rsr and shown that Rsr binds Y RNA with nanomolar affinity. The high resolution crystal structure of Rsr when compared to the structure of *X. laevis* Ro suggest obvious structural changes in the Ro protein. These conformational changes suggest that Ro proteins possesses a high degree of flexibility at the interface of the two domains. Re-arrangement of helices at this interface results in alterations in the size of the central RNA binding cavity. This high degree of flexibility may be possible due to conserved Glycine-rich motifs as seen in other HEAT repeat proteins. The structural flexibility of Ro may be required to adapt to the different RNA substrates that are bound at the central cavity, and possibly for facilitating the entry and release of RNA substrates. Our hypothesis based on the crystal structure, needs to be tested. To test this hypothesis, and show beyond doubt that the opening and closing of the central cavity is governed by the H15 to H18 α-helices, many solution-state experimental approaches could be used. Fluorescence based assays could be used to report on local movements near the α-helices. Also, disulphide cross-linking experiments have proven particularly useful in observing local conformational changes. This would involve introducing cysteine residues at critical places in Rsr, such that only certain conformations give rise to disulfide cross-links. In the presence and absence of the disulfide cross-links, the RNA binding activity would have to be tested.

Crystallography, though a powerful tool to obtain structural detail to an atomic resolution, provides a snap-shot or a static image of the macromolecule. To gain detailed mechanistic insights, a series of such snap-shots are required and very often, even these may not provide a complete picture. In such cases, a combination of crystallography and solution studies could prove to be more informative. The structure of Rsr provides a solid foundation for future work involving structure based probing of Ro RNA interactions and its mechanistic details.

While a wealth of information is available for Ro in eukaryotes, relatively less in known about the biological role of Ro in prokaryotes. The structure of Rsr provides clues to its function. Our results show that structurally the Ro proteins from prokaryotes are similar to eukaryotic Ro. Based on this, we expect Ro to be involved in binding misfolded or mutant RNAs just like it has been shown for eukaryotic Ro proteins. Whether or not Ro is directly involved in the folding of these RNAs is not known, and future research needs to address this question. Interactions between Ro and RNA processing enzymes like helicases, nucleases etc. need to be extensively studied to understand the role of Ro. Alternatively, it is possible that Ro simply sequesters misfolded RNAs to prevent them from assembling into cellular machinery, thereby acting as a quality control protein. Identification of cognate RNAs that bind at the central cavity of Rsr would greatly improve our understanding of the function of Ro in prokaryotes. Preliminary crystal trials using single stranded RNAs that could possibly bind Rsr in the central cavity have proven unsuccessful. Identification of the biologically relevant RNAs that bind Rsr in the central cavity would greatly aid not only in the crystallization experiments, but also in the overall understanding of prokaryotic Ro proteins.

Our finding that Rsr interacts with Y RNA to form large multimers is intriguing. The physiological relevance of this observation needs to be investigated. Previous work on human Ro RNPs suggests that Ro RNPs may exist as large (~350kDa) complexes in the cell. In the cell, the composition of these complexes is not clear yet, since other proteins could be associated with the Ro RNPs. Our *in vitro* experiments provide clear evidence that Ro and Y RNA could associate to form large oligomers in the absence of other factors. It is possible that formation of these large oligomers within the cell acts as a trigger for the recognition of this complex by the yet unidentified downstream processing events. Research over the last few decades has helped identify Ro RNPs as being the primary antigenic species in many autoimmune disorders. However, even today, it is not known what switches Ro RNPs from being normal cellular components to what the immune system recognizes as “foreign”. It is possible that oligomerization

under certain conditions in the cell, could determine this switch from normal cellular components to “foreign”. Research in this direction could provide valuable clues towards understanding lupus-like auto-immune disorders.

Collectively, our structural studies have now provided a strong foundation for future research on Ro proteins and the experiments proposed should facilitate better understanding of the function of this unique protein, in both prokaryotes and eukaryotes.

4.2 STRUCTURAL STUDY OF CsoR FROM *M. tuberculosis*

In this work, we have determined a high resolution structure of Cu(I) bound CsoR. The structure shows that the metal site forms at the dimer interface and involves ligands from both protomers. Research on other metalloregulators has shown that the nature of the ligands presented at the metal site, the geometry of the metal site and consequently, the chemical environment at the metal site are all determinants of metal specificity in these proteins. The structure of CsoR suggests that metal specificity of CsoR may be determined in a similar way. CsoR presents ligands that are typically preferred by Cu(I), like cysteines and histidines, in a geometry that is favorable for Cu(I) binding. In this way, CsoR is likely to distinguish not only between copper and other metals, but also between different oxidation states of copper itself. Further understanding of CsoR’s metal binding properties could some from structural studies on the metal-free state of CsoR. Determining the structure of metal-free CsoR using crystallography has proven to be a challenge. Although the protein forms crystals, the crystals are not of high enough quality for structure determination. Another problem is the formation of isomorphous crystals for Apo and Cu(I) bound CsoR. This may be due to trace amounts of metal bound to the Apo protein, or a result of the crystallization process, where the protein may be trapped in a certain conformation. Future research needs to address these issues. An approach that combines crystallography with solution state techniques like NMR and fluorescence, could provide valuable information about the structural changes that CsoR undergoes upon Cu(I) binding.

Comparison of the Cu(I):CsoR structure with previously characterized Cu(I)

responsive transcriptional regulators reveals obvious differences in the mode of DNA binding. Based on the crystal structure, CsoR does not seem to possess the typical DNA-binding winged helix or helix-turn-helix motifs seen in other metal responsive transcriptional regulators. This suggests that CsoR is the founding member of a new structural class of metalloregulators. DALI searches for structurally similar proteins did not yield positive results, suggesting that CsoR may possess a novel DNA binding fold. To understand the basis of DNA recognition by CsoR, we have made attempts to crystallize the CsoR:DNA complex. These have proven to be unsuccessful since the DNA tends to crystallize on its own. Of the many metalloregulators that have been characterized, high resolution structures in complex with the operator/promoter DNA are available only for three proteins. All three of these belong to the DtxR and NikR families that bind DNA in their metal-loaded forms. Using crystallography to obtain structures for DNA complexes of metalloregulators that bind DNA in their apo form (like CsoR) may be even more challenging since crystal screening involves many hundred conditions often with trace amounts of metal in them. In this respect, other solution based methods like NMR may prove to be more successful for such studies.

As a Cu(I) responsive regulator of transcription, CsoR must not only bind Cu(I), but also modulate its interactions with DNA upon sensing Cu(I). The mechanism that governs the switch between the two states is all important for CsoR function. Our work provides clues to the mechanism. In the structure, close to the Cu(I) site, a conserved glutamate residue is present which can possibly mediate interactions between the Cu(I) site and the possible DNA binding surface. Future work needs to be directed towards understanding the mechanistic details of the switch in CsoR from DNA bound to Cu(I) bound states. Structure-based mutational analysis of the protein, along with other biochemical techniques to study protein-DNA and protein-metal interactions could provide greater mechanistic insights.

Preliminary characterization of the *cso* operon suggests an elegant sequence of events when cells are subjected to an excess of Cu(I). CsoR responds to free Cu(I) in the cell, causing derepression of the *cso* operon. The genes in the *cso* operon, whose

expression is regulated by CsoR, are predicted to encode a hypothetical conserved protein (*rv 0968)* and a putative Cu(I) translocating P-type ATPase (*rv 0969).* A detailed understanding of the cellular response to Cu(I) would require a thorough characterization of the gene products of the *cso* operon. This would help is understand the steps that take place after the derepression by CsoR. It is possible that CsoR interacts with either of the two proteins (Rv0968 and Rv0969) to deliver Cu(I) to them, or that the conserved hypothetical protein (Rv0968) may act as a Cu(I) chaperone, mediating copper delivery to the P-type ATPase.

Research on *Mycobacterium tuberculosis* in the last few decades has shown beyond doubt, the importance of metals like iron, zinc and copper (197). However, our understanding of copper homeostasis in *M. tb* in the macrophage environment is still in the early stages. The discovery of CsoR as the primary component of copper homeostasis in *M. tb* and many other pathogenic bacteria, along with the structural aspects of Cu(I) binding by CsoR provide a good starting point for future research towards understanding copper homeostasis in intracellular pathogens.

REFERENCES

1. Pickles, B., Jeffrey, B. A., and Rossman, M. G. (1964) *J Mol Biol* **93**, 598-600
2. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) *Nature* **181**(4610), 662-666
3. Watson, J. D., and Crick, F. H. (1953) *Cold Spring Harb Symp Quant Biol* **18**, 123-131
4. Alberts, B. (1994) *Molecular biology of the cell*, 3 Ed., Garland publishing Inc., New York.
5. Lin, H. (2007) *Science* **316**(5823), 397
6. Carl Brandon, J. T. (1998) *Introduction to protein structure*, 2 Ed., Garland Publishing Inc., New York.
7. Garvie, C. W., and Wolberger, C. (2001) *Mol Cell* **8**(5), 937-946
8. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) *Nucleic Acids Res* **29**(13), 2860-2874
9. Beamer, L. J., and Pabo, C. O. (1992) *J Mol Biol* **227**(1), 177-196
10. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) *Cell* **67**(3), 517-528
11. Suzuki, M., and Gerstein, M. (1995) *Proteins* **23**(4), 525-535
12. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) *Cell* **78**(2), 211-223
13. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L., and Wolberger, C. (1998) *Science* **279**(5353), 1037-1041
14. Changela, A., Chen, K., Xue, Y., Holschen, J., Outten, C. E., O'Halloran, T. V., and Mondragon, A. (2003) *Science* **301**(5638), 1383-1387
15. Wisedchaisri, G., Chou, C. J., Wu, M., Roach, C., Rice, A. E., Holmes, R. K., Beeson, C., and Hol, W. G. (2007) *Biochemistry* **46**(2), 436-447
16. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science* **254**(5031),

539-544

17. Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature* **363**(6424), 38-45
18. Bell, C. E., and Lewis, M. (2000) *Nat Struct Biol* **7**(3), 209-214
19. Schumacher, M. A., Choi, K. Y., Zalkin, H., and Brennan, R. G. (1994) *Science* **266**(5186), 763-770
20. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure* **4**(10), 1171-1180
21. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S. C. (1992) *Nature* **356**(6368), 408-414
22. Weihofen, W. A., Cicek, A., Pratto, F., Alonso, J. C., and Saenger, W. (2006) *Nucleic Acids Res* **34**(5), 1450-1458
23. Schreiter, E. R., Wang, S. C., Zamble, D. B., and Drennan, C. L. (2006) *Proc Natl Acad Sci U S A* **103**(37), 13676-13681
24. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) *Nature* **367**(6465), 754-757
25. Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) *Nature* **365**(6446), 512-520
26. Cramer, P., Larson, C. J., Verdine, G. L., and Muller, C. W. (1997) *Embo J* **16**(23), 7078-7090
27. Holbrook, S. R. (2005) *Curr Opin Struct Biol* **15**(3), 302-308
28. Hendrix, D. K., Brenner, S. E., and Holbrook, S. R. (2005) *Q Rev Biophys* **38**(3), 221-243
29. Saenger, W. (1984) *Principles of nucleic-acid structure*, Springer-Verlag New York Inc., New York.
30. Holbrook, S. R., Cheong, C., Tinoco, I., Jr., and Kim, S. H. (1991) *Nature* **353**(6344), 579-581
31. White, S. A., Nilges, M., Huang, A., Brunger, A. T., and Moore, P. B. (1992) *Biochemistry* **31**(6), 1610-1621

32. Perez-Canadillas, J. M., and Varani, G. (2001) *Curr Opin Struct Biol* **11**(1), 53-58
33. Willingham, A. T., and Gingeras, T. R. (2006) *Cell* **125**(7), 1215-1220
34. Spirin, A. S. (2004) *RNA Biol* **1**(1), 3-9
35. Finnegan, E. J., and Matzke, M. A. (2003) *J Cell Sci* **116**(Pt 23), 4689-4693
36. Hall, K. B. (2002) *Curr Opin Struct Biol* **12**(3), 283-288
37. Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M., and Thornton, J. M. (2001) *Nucleic Acids Res* **29**(4), 943-954
38. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990) *Nature* **348**(6301), 515-520
39. Price, S. R., Evans, P. R., and Nagai, K. (1998) *Nature* **394**(6694), 645-650
40. Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) *Nucleic Acids Res* **21**(5), 1193-1198
41. Backe, P. H., Messias, A. C., Ravelli, R. B., Sattler, M., and Cusack, S. (2005) *Structure* **13**(7), 1055-1067
42. Backe, P. H., Ravelli, R. B., Garman, E., and Cusack, S. (2004) *Acta Crystallogr D Biol Crystallogr* **60**(Pt 4), 784-787
43. Wang, X., McLachlan, J., Zamore, P. D., and Hall, T. M. (2002) *Cell* **110**(4), 501-512
44. Bateman, A., and Kickhoefer, V. (2003) *BMC Bioinformatics* **4**, 49
45. Fierro-Monti, I., and Mathews, M. B. (2000) *Trends Biochem Sci* **25**(5), 241-246
46. Saunders, L. R., and Barber, G. N. (2003) *FASEB J* **17**(9), 961-983
47. Ryter, J. M., and Schultz, S. C. (1998) *EMBO J* **17**(24), 7505-7513
48. Ramos, A., Grunert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000) *EMBO J* **19**(5), 997-1009
49. Wolin, S. L., and Steitz, J. A. (1984) *Proc Natl Acad Sci U S A* **81**(7), 1996-2000
50. Wolin, S. L., and Steitz, J. A. (1983) *Cell* **32**(3), 735-744
51. Mattioli, M., and Reichlin, M. (1974) *Arthritis Rheum* **17**(4), 421-429
52. Lerner, M. R., Boyle, J. A., Hardin, J. A., and Steitz, J. A. (1981) *Science*

211(4480), 400-402

53. Alspaugh, M. A., and Tan, E. M. (1975) *J Clin Invest* **55**(5), 1067-1073
54. Moldovan, I. (2006) *Compr Ther* **32**(3), 158-162
55. Lee, L. A., Alvarez, K., Gross, T., and Harley, J. B. (1996) *J Invest Dermatol* **107**(2), 225-228
56. Boire, G., Lopez-Longo, F. J., Lapointe, S., and Menard, H. A. (1991) *Arthritis Rheum* **34**(6), 722-730
57. Alarcon-Segovia, D. (2001) *Isr Med Assoc J* **3**(2), 127-130
58. Mamula, M. J., O'Brien, C. A., Harley, J. B., and Hardin, J. A. (1989) *Clin Immunol Immunopathol* **52**(3), 435-446
59. Van Horn, D. J., Eisenberg, D., O'Brien, C. A., and Wolin, S. L. (1995) *RNA* **1**(3), 293-303
60. O'Brien, C. A., Margelot, K., and Wolin, S. L. (1993) *Proc Natl Acad Sci U S A* **90**(15), 7250-7254
61. Hendrick, J. P., Wolin, S. L., Rinke, J., Lerner, M. R., and Steitz, J. A. (1981) *Mol Cell Biol* **1**(12), 1138-1149
62. Christov, C. P., Gardiner, T. J., Szuts, D., and Krude, T. (2006) *Mol Cell Biol* **26**(18), 6993-7004
63. Wolin, S. L., and Cedervall, T. (2002) *Annu Rev Biochem* **71**, 375-403
64. Ben-Chetrit, E., Chan, E. K., Sullivan, K. F., and Tan, E. M. (1988) *J Exp Med* **167**(5), 1560-1571
65. Kelekar, A., Saitta, M. R., and Keene, J. D. (1994) *J Clin Invest* **93**(4), 1637-1644
66. Wahren, M., Mellqvist, E., Vene, S., Ringertz, N. R., and Pettersson, I. (1996) *Eur J Cell Biol* **70**(3), 189-197
67. Xue, D., Shi, H., Smith, J. D., Chen, X., Noe, D. A., Cedervall, T., Yang, D. D., Eynon, E., Brash, D. E., Kashgarian, M., Flavell, R. A., and Wolin, S. L. (2003) *Proc. Natl. Acad. Sci. U S A* **100**(13), 7503-7508
68. Chen, X., Smith, J. D., Shi, H., Yang, D. D., Flavell, R. A., and Wolin, S. L.

(2003) *Curr Biol* **13**(24), 2206-2211

69. O'Brien, C. A., and Wolin, S. L. (1994) *Genes Dev* **8**(23), 2891-2903
70. Huber, P. W., Rife, J. P., and Moore, P. B. (2001) *J Mol Biol* **312**(4), 823-832
71. Shi, H., O'Brien, C. A., Van Horn, D. J., and Wolin, S. L. (1996) *RNA* **2**(8), 769-784
72. Fuchs, G., Stein, A. J., Fu, C., Reinisch, K. M., and Wolin, S. L. (2006) *Nat Struct Mol Biol*
73. Barciszewska, M. Z., Szymanski, M., Erdmann, V. A., and Barciszewski, J. (2001) *Acta Biochim Pol* **48**(1), 191-198
74. Xing, Y. Y., and Worcel, A. (1989) *Genes Dev* **3**(7), 1008-1018
75. Picard, B., and Wegnez, M. (1979) *Proc Natl Acad Sci U S A* **76**(1), 241-245
76. Rudt, F., and Pieler, T. (1996) *Embo J* **15**(6), 1383-1391
77. Honda, B. M., and Roeder, R. G. (1980) *Cell* **22**(1 Pt 1), 119-126
78. Joho, K. E., Darby, M. K., Crawford, E. T., and Brown, D. D. (1990) *Cell* **61**(2), 293-300
79. Chen, X., and Wolin, S. L. (2004) *J Mol Med* **82**(4), 232-239
80. Glockner, F. O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R., Schlesner, H., Amann, R., and Reinhardt, R. (2003) *Proc Natl Acad Sci U S A* **100**(14), 8298-8303
81. Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J. G., Jacobs, W. R., Jr., Hendrix, R. W., and Hatfull, G. F. (2003) *Cell* **113**(2), 171-182
82. Cox, M. M., and Battista, J. R. (2005) *Nat Rev Microbiol* **3**(11), 882-892
83. Chen, X., Quinn, A. M., and Wolin, S. L. (2000) *Genes Dev* **14**(7), 777-782
84. Levin-Zaidman, S., Englander, J., Shimoni, E., Sharma, A. K., Minton, K. W., and Minsky, A. (2003) *Science* **299**(5604), 254-256
85. Archibald, F. S., and Fridovich, I. (1981) *J Bacteriol* **146**(3), 928-936

86. Minton, K. W. (1994) *Mol Microbiol* **13**(1), 9-15
87. Daly, M. J., and Minton, K. W. (1996) *J Bacteriol* **178**(15), 4461-4471
88. Narumi, I., Satoh, K., Cui, S., Funayama, T., Kitayama, S., and Watanabe, H. (2004) *Mol Microbiol* **54**(1), 278-285
89. Green, C. D., Long, K. S., Shi, H., and Wolin, S. L. (1998) *RNA* **4**(7), 750-765
90. Stein, A. J., Fuchs, G., Fu, C., Wolin, S. L., and Reinisch, K. M. (2005) *Cell* **121**(4), 529-539
91. Andrade, M. A., and Bork, P. (1995) *Nat Genet* **11**(2), 115-116
92. Tuckwell, D. (1999) *Biochem Soc Trans* **27**(6), 835-840
93. Whittaker, C. A., and Hynes, R. O. (2002) *Mol Biol Cell* **13**(10), 3369-3387
94. Bergelson, J. M., and Hemler, M. E. (1995) *Curr Biol* **5**(6), 615-617
95. Ramesh, A., Savva, C. G., Holzenburg, A., and Sacchettini, J. C. (2007) *J Biol Chem*
96. Bandziulis, R. J., Swanson, M. S., and Dreyfuss, G. (1989) *Genes Dev* **3**(4), 431-437
97. Kenan, D. J., Query, C. C., and Keene, J. D. (1991) *Trends Biochem Sci* **16**(6), 214-220
98. Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993) *Annu Rev Biochem* **62**, 289-321
99. O. Lund, M. N., Lundegaard, C., Worning, P. (2002) CPHmodels 2.0: X3M a Computer Program to Extract 3D Models. In. *Abstract at the CASP5 ConferenceA102*
100. Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998) *Cell* **94**(2), 193-204
101. Lee, S. J., Sekimoto, T., Yamashita, E., Nagoshi, E., Nakagawa, A., Imamoto, N., Yoshimura, M., Sakai, H., Chong, K. T., Tsukihara, T., and Yoneda, Y. (2003) *Science* **302**(5650), 1571-1575
102. Lee, J. O., Bankston, L. A., Arnaout, M. A., and Liddington, R. C. (1995) *Structure* **3**(12), 1333-1340

103. Lacy, D. B., Wigelsworth, D. J., Scobie, H. M., Young, J. A., and Collier, R. J. (2004) *Proc Natl Acad Sci U S A* **101**(17), 6367-6372

104. Lee, J. O., Rieu, P., Arnaout, M. A., and Liddington, R. (1995) *Cell* **80**(4), 631-638

105. Shimaoka, M., Takagi, J., and Springer, T. A. (2002) *Annu Rev Biophys Biomol Struct* **31**, 485-516

106. Boire, G., and Craft, J. (1990) *J Clin Invest* **85**(4), 1182-1190

107. Zachariae, U., and Grubmuller, H. (2006) *Structure* **14**(9), 1469-1478

108. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1), 2256-2268

109. Jacob, J., Duclohier, H., and Cafiso, D. S. (1999) *Biophys J* **76**(3), 1367-1376

110. Rosenhouse-Dantsker, A., and Logothetis, D. E. (2006) *Biophys J* **91**(8), 2860-2873

111. Farris, A. D., Koelsch, G., Pruijn, G. J., van Venrooij, W. J., and Harley, J. B. (1999) *Nucleic Acids Res* **27**(4), 1070-1078

112. Conti, E., Muller, C. W., and Stewart, M. (2006) *Curr Opin Struct Biol* **16**(2), 237-244

113. Muzzolini, L., Beuron, F., Patwardhan, A., Popuri, V., Cui, S., Niccolini, B., Rappas, M., Freemont, P. S., and Vindigni, A. (2007) *PLoS Biol* **5**(2), e20

114. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci* **4**(11), 2411-2423

115. McPherson, A. (1991) *Crystallization of biological macromolecules*, 1 Ed., Cold Spring Harbor Laboratory Press

116. Rhodes, G. (2000) *Crystallography made crystal clear*, 2 Ed., Academic Press, San Diego, California

117. Blow, D. (2002) *Outline of crystallography for biologists*, Oxford University Press, Oxford

118. McRee, D. E. (1999) *Practical Protein Crystallography*, 2 Ed., Academic Press, San Diego , California

119. *International tables of crystallography*, Kluwer Academic Publisher, Netherlands
120. Ioerger, T. R., and Sacchettini, J. C. (2003) *Methods Enzymol* **374**, 244-270
121. Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., and Terwilliger, T. C. (2004) *J Synchrotron Radiat* **11**(Pt 1), 53-55
122. Minor, Z. O. a. W. (1997) *Methods in Enzymology, Macromolecular Crystallography, part A.* **276**, 307-326
123. Cowtan, K. D., and Main, P. (1996) *Acta Crystallogr D Biol Crystallogr* **52**(Pt 1), 43-48
124. McRee, D. E. (1999) *J Struct Biol* **125**(2-3), 156-165
125. Reddy, V., Swanson, S. M., Segelke, B., Kantardjieff, K. A., Sacchettini, J. C., and Rupp, B. (2003) *Acta Crystallogr D Biol Crystallogr* **59**(Pt 12), 2200-2210
126. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54**(Pt 5), 905-921
127. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc Natl Acad Sci U S A* **98**(18), 10037-10041
128. Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996) *Biopolymers* **38**(3), 305-320
129. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J Comput Chem* **25**(13), 1605-1612
130. Riley-Lovingshimer, M. R., and Reinhart, G. D. (2005) *Arch Biochem Biophys* **436**(1), 178-186
131. Valentine, R. C., Shapiro, B. M., and Stadtman, E. R. (1968) *Biochemistry* **7**(6), 2143-2152
132. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) *J Struct Biol* **128**(1), 82-97
133. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996) *J*

Struct Biol **116**(1), 17-24

134. Van Heel, M. (1989) *Optik* **82**, 114- 126
135. Fraustoda Silva J, W. R. (2001) *The biological chemistry of elements: The inorganic chemistry of life,* Oxford University Press, Oxford
136. Carr, H. S., and Winge, D. R. (2003) *Acc Chem Res* **36**(5), 309-316
137. Cobine, P. A., Pierrel, F., and Winge, D. R. (2006) *Biochim Biophys Acta* **1763**(7), 759-772
138. Michel, H., Behr, J., Harrenga, A., and Kannt, A. (1998) *Annu Rev Biophys Biomol Struct* **27**, 329-356
139. Fridovich, I. (1978) *Science* **201**(4359), 875-880
140. Fridovich, I. (1978) *Ciba Found Symp* (65), 77-93
141. Brazeau, B. J., Johnson, B. J., and Wilmot, C. M. (2004) *Arch Biochem Biophys* **428**(1), 22-31
142. Mure, M. (2004) *Acc Chem Res* **37**(2), 131-139
143. Bremner, I. (1998) *Am J Clin Nutr* **67**(5 Suppl), 1069S-1073S
144. Kadiiska, M. B., Hanna, P. M., Jordan, S. J., and Mason, R. P. (1993) *Mol Pharmacol* **44**(1), 222-227
145. Powell, S. R. (2000) *J Nutr* **130**(5S Suppl), 1447S-1454S
146. Buettner, G. R. (1993) *Arch Biochem Biophys* **300**(2), 535-543
147. Kawanishi, S., Inoue, S., and Yamamoto, K. (1989) *Biol Trace Elem Res* **21**, 367-372
148. O'Connor, J. M. (2001) *Biochem Soc Trans* **29**(Pt 2), 354-357
149. Liang, Q., and Dedon, P. C. (2001) *Chem Res Toxicol* **14**(4), 416-422
150. Chow, C. K. (1979) *Am J Clin Nutr* **32**(5), 1066-1081
151. Requena, J. R., Groth, D., Legname, G., Stadtman, E. R., Prusiner, S. B., and Levine, R. L. (2001) *Proc Natl Acad Sci U S A* **98**(13), 7170-7175
152. Inestrosa, N. C., Cerpa, W., and Varela-Nallar, L. (2005) *IUBMB Life* **57**(9), 645-650
153. Rasia, R. M., Bertoncini, C. W., Marsh, D., Hoyer, W., Cherny, D., Zweckstetter,

M., Griesinger, C., Jovin, T. M., and Fernandez, C. O. (2005) *Proc Natl Acad Sci U S A* **102**(12), 4294-4299

154. Rossi, L., Arciello, M., Capo, C., and Rotilio, G. (2006) *Ital J Biochem* **55**(3-4), 212-221
155. Eakin, C. M., Berman, A. J., and Miranker, A. D. (2006) *Nat Struct Mol Biol* **13**(3), 202-208
156. Dameron, C. T., and Harrison, M. D. (1998) *Am J Clin Nutr* **67**(5 Suppl), 1091S-1097S
157. Hassett, R., Dix, D. R., Eide, D. J., and Kosman, D. J. (2000) *Biochem J* **351 Pt 2**, 477-484
158. Nose, Y., Rees, E. M., and Thiele, D. J. (2006) *Trends Biochem Sci* **31**(11), 604-607
159. Wunderli-Ye, H., and Solioz, M. (2001) *Biochem Biophys Res Commun* **280**(3), 713-719
160. Vulpe, C. D., and Packman, S. (1995) *Annu Rev Nutr* **15**, 293-322
161. Palmgren, M. G., and Axelsen, K. B. (1998) *Biochim Biophys Acta* **1365**(1-2), 37-45
162. Vulpe, C., Levinson, B., Whitney, S., Packman, S., and Gitschier, J. (1993) *Nat Genet* **3**(1), 7-13
163. Tottey, S., Rich, P. R., Rondet, S. A., and Robinson, N. J. (2001) *J Biol Chem* **276**(23), 19999-20004
164. Kanamaru, K., Kashiwagi, S., and Mizuno, T. (1994) *Mol Microbiol* **13**(2), 369-377
165. Kille, P., Hemmings, A., and Lunney, E. A. (1994) *Biochim Biophys Acta* **1205**(2), 151-161
166. Grill, E., Loffler, S., Winnacker, E. L., and Zenk, M. H. (1989) *Proc Natl Acad Sci U S A* **86**(18), 6838-6842
167. Thumann, J., Grill, E., Winnacker, E. L., and Zenk, M. H. (1991) *FEBS Lett* **284**(1), 66-69

168. Rosenzweig, A. C., and O'Halloran, T. V. (2000) *Curr Opin Chem Biol* **4**(2), 140-147

169. Culotta, V. C., Klomp, L. W., Strain, J., Casareno, R. L., Krems, B., and Gitlin, J. D. (1997) *J Biol Chem* **272**(38), 23469-23472

170. Pufahl, R. A., Singer, C. P., Peariso, K. L., Lin, S. J., Schmidt, P. J., Fahrni, C. J., Culotta, V. C., Penner-Hahn, J. E., and O'Halloran, T. V. (1997) *Science* **278**(5339), 853-856

171. Lamb, A. L., Wernimont, A. K., Pufahl, R. A., Culotta, V. C., O'Halloran, T. V., and Rosenzweig, A. C. (1999) *Nat Struct Biol* **6**(8), 724-729

172. Banci, L., Bertini, I., Ciofi-Baffoni, S., Del Conte, R., and Gonnelli, L. (2003) *Biochemistry* **42**(7), 1939-1949

173. Arnesano, F., Banci, L., Bertini, I., Cantini, F., Ciofi-Baffoni, S., Huffman, D. L., and O'Halloran, T. V. (2001) *J Biol Chem* **276**(44), 41365-41376

174. Pennella, M. A., and Giedroc, D. P. (2005) *Biometals* **18**(4), 413-428

175. Wunderli-Ye, H., and Solioz, M. (1999) *Biochem Biophys Res Commun* **259**(2), 443-449

176. Chen, K., Yuldasheva, S., Penner-Hahn, J. E., and O'Halloran, T. V. (2003) *J Am Chem Soc* **125**(40), 12088-12089

177. Busenlehner, L. S., Pennella, M. A., and Giedroc, D. P. (2003) *FEMS Microbiol Rev* **27**(2-3), 131-143

178. Liu, T., Nakashima, S., Hirose, K., Shibasaka, M., Katsuhara, M., Ezaki, B., Giedroc, D. P., and Kasamo, K. (2004) *J Biol Chem* **279**(17), 17810-17818

179. Wu, J., and Rosen, B. P. (1993) *J Biol Chem* **268**(1), 52-58

180. Morby, A. P., Turner, J. S., Huckle, J. W., and Robinson, N. J. (1993) *Nucleic Acids Res* **21**(4), 921-925

181. Chou, C. J., Wisedchaisri, G., Monfeli, R. R., Oram, D. M., Holmes, R. K., Hol, W. G., and Beeson, C. (2004) *J Biol Chem* **279**(51), 53554-53561

182. Andrews, S. C., Robinson, A. K., and Rodriguez-Quinones, F. (2003) *FEMS Microbiol Rev* **27**(2-3), 215-237

183. Strausak, D., and Solioz, M. (1997) *J Biol Chem* **272**(14), 8932-8936
184. White, A., Ding, X., vanderSpek, J. C., Murphy, J. R., and Ringe, D. (1998) *Nature* **394**(6692), 502-506
185. Dameron, C. T., Winge, D. R., George, G. N., Sansone, M., Hu, S., and Hamer, D. (1991) *Proc Natl Acad Sci U S A* **88**(14), 6127-6131
186. Brown, K. R., Keller, G. L., Pickering, I. J., Harris, H. H., George, G. N., and Winge, D. R. (2002) *Biochemistry* **41**(20), 6469-6476
187. Clemens, S., Kim, E. J., Neumann, D., and Schroeder, J. I. (1999) *Embo J* **18**(12), 3325-3333
188. Solioz, M., Odermatt, A., and Krapf, R. (1994) *FEBS Lett* **346**(1), 44-47
189. Petris, M. J., Smith, K., Lee, J., and Thiele, D. J. (2003) *J Biol Chem* **278**(11), 9639-9646
190. Tottey, S., Harvie, D. R., and Robinson, N. J. (2005) *Acc Chem Res* **38**(10), 775-783
191. Hullo, M. F., Moszer, I., Danchin, A., and Martin-Verstraete, I. (2001) *J Bacteriol* **183**(18), 5426-5430
192. Cavet, J. S., Borrelly, G. P., and Robinson, N. J. (2003) *FEMS Microbiol Rev* **27**(2-3), 165-181
193. Wandersman, C., and Delepelaire, P. (2004) *Annu Rev Microbiol* **58**, 611-647
194. Flo, T. H., Smith, K. D., Sato, S., Rodriguez, D. J., Holmes, M. A., Strong, R. K., Akira, S., and Aderem, A. (2004) *Nature* **432**(7019), 917-921
195. Mazmanian, S. K., Skaar, E. P., Gaspar, A. H., Humayun, M., Gornicki, P., Jelenska, J., Joachmiak, A., Missiakas, D. M., and Schneewind, O. (2003) *Science* **299**(5608), 906-909
196. Francis, M. S., and Thomas, C. J. (1997) *Microb Pathog* **22**(2), 67-78
197. Agranoff, D. (2004) *Frontiers in Bioscience* **9**, 2996-3006
198. Wu, C. H., Tsai-Wu, J. J., Huang, Y. T., Lin, C. Y., Lioua, G. G., and Lee, F. J. (1998) *FEBS Lett* **439**(1-2), 192-196
199. Gobin, J., Moore, C. H., Reeve, J. R., Jr., Wong, D. K., Gibson, B. W., and

Horwitz, M. A. (1995) *Proc Natl Acad Sci U S A* **92**(11), 5189-5193

200. Pym, A. S., Domenech, P., Honore, N., Song, J., Deretic, V., and Cole, S. T. (2001) *Mol Microbiol* **40**(4), 879-889
201. Gold, B., Rodriguez, G. M., Marras, S. A., Pentecost, M., and Smith, I. (2001) *Mol Microbiol* **42**(3), 851-865
202. Wang, Y., Hemmingsen, L., and Giedroc, D. P. (2005) *Biochemistry* **44**(25), 8976-8988
203. Cavet, J. S., Meng, W., Pennella, M. A., Appelhoff, R. J., Giedroc, D. P., and Robinson, N. J. (2002) *J Biol Chem* **277**(41), 38441-38448
204. Iwig, J. S., Rowe, J. L., and Chivers, P. T. (2006) *Mol Microbiol* **62**(1), 252-262
205. Eicken, C., Pennella, M. A., Chen, X., Koshlap, K. M., VanZile, M. L., Sacchettini, J. C., and Giedroc, D. P. (2003) *J Mol Biol* **333**(4), 683-695
206. Gajiwala, K. S., and Burley, S. K. (2000) *Curr Opin Struct Biol* **10**(1), 110-116
207. Furey, W. (1997) *Methods in Enzymology* **277**(31)**,** 590-620
208. Ioerger, T. R., and Sacchettini, J. C. (2002) *Acta Crystallogr D Biol Crystallogr* **58**(Pt 12), 2043-2054
209. (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 760-763
210. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-921
211. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**(15), 2714-2723

VITA

Name: Arati Ramesh

Address: 17/1C 'Raghuveda', Jagadeeswaran St., T-Nagar, Chennai 600017, India.

Email address: arati.ramesh@gmail.com

Education: B.S. Genetics, St. Ann's College, Hyderabad, India. 1998

M.S. Biotechnology, University of Pune, Pune, India. 2000

Employment: Unilever Research Intl., Bangalore, India. 2000-2001